Psychological Review
1990, Vol. 97, No. 2, 201–224

0033-295X/90/$00.75

Introduction to Knowledge Spaces: How to Build, Test, and Search Them

Jean-Claude Falmagne
University of California, Irvine

Jean-Paul Doignon
Université Libre de Bruxelles
Brussels, Belgium

Mathieu Koppen and Michael Villano
New York University

Leila Johannesen
Honeywell Systems and Research Center
Minneapolis, Minnesota

This article gives a comprehensive description of a theory for the efficient assessment of knowledge. The essential concept is that the *knowledge state* of a subject with regard to a specified field of information can be represented by a particular subset of questions or problems that the subject is capable of solving. The family of all knowledge states forms the *knowledge space.* It is assumed that if 2 subsets K and K' of questions are assumed to be states in a knowledge space $\mathcal{K}$, then $K \cup K'$ is also assumed to be a state in $\mathcal{K}$. Such a theory is consistent with the idea that at least some of the notions in the field may be acquired from different sets of prerequisites. Various aspects of the theory are discussed. In particular, the problem of constructing a knowledge space in practice is analyzed in detail. A first sketch of the knowledge space can be obtained by consulting expert teachers in the field. The mathematical theory necessary to render this consultation efficient is given. This preliminary construction can then be tested and refined on the basis of empirical data. To this end, a probabilistic version of the theory is developed, which is similar in spirit to some psychometric models, but it is grounded on the concept of a knowledge space rather than on that of skill or ability. An exemplary application of this probabilistic theory to a high school mathematics test is described, based on a sample of several hundred students. By standard likelihood ratio methods, it is shown how the preliminary knowledge space can be gradually refined, and the number of possible knowledge states substantially reduced. Two classes of Markovian knowledge assessment algorithms are outlined. Most of the results presented summarize previous articles published in various technical journals. The application of the probabilistic theory to the high school mathematics test is original to this article.

Consider the task of assessing an individual's knowledge state with regard to a specified collection of problems or questions. For concreteness, a small sample of such problems is shown below, taken from the standard high school mathematics curriculum. This example will be used throughout the article.

1. $\begin{array}{r} 378 \\ \underline{\times 605} \\ ? \end{array}$
2. $58.7 \times 0.94 = ?$
3. $\frac{1}{2} \times \frac{5}{6} = ?$
4. What is 30% of 34?
5. Gwendolyn is $\frac{3}{4}$ as old as Rebecca.

   Rebecca is $\frac{2}{5}$ as old as Edwin.

   Edwin is 20 years old. How old is Gwendolyn?

Testers could ask all five problems and examine the responses. However, such a method would be inefficient and impractical if they had 10,000 students to examine on an 80-problem test. Efficiency can be improved by taking advantage of an implicit structure relating these problems. For instance, we can safely deduce that any student capable of solving Problem 5 can also solve Problems 1, 2, and 3 (we shall verify this experimentally later in the article). Thus, asking those three problems would be a waste of time. Similarly, a student who fails Problem 1 should almost certainly not be asked Problems 2, 4, or 5.

Before proceeding, let us eliminate some possible misunderstandings. We are not assuming that the structure is necessarily

The work reported herein was supported by National Science Foundation Grant IST-8418860 and Army Research Institute Grant DAAG29-84-G0083 to Jean-Claude Falmagne at New York University. The article was written while he was a fellow at the Center for Advanced Study in the Behavioral Sciences. The financial support of the National Science Foundation (BNS-870086) and of the Alfred P. Sloan Foundation is gratefully acknowledged. Jean-Paul Doignon's contribution was made possible by a Fulbright grant and a North Atlantic Treaty Organization scientific grant.

We thank Donald Bamber, Bill Batchelder, and Duncan Luce for their useful reactions on drafts of this article. Finally, we are indebted to Joey Gordon for serving as an expert teacher and providing the basic data for the initial determination of the knowledge space and to Maria Kambouri for helping us with the collection and analysis of the student data.

Correspondence concerning this article should be addressed to Jean-Claude Falmagne, School of Social Sciences, University of California, Irvine, California 92717.

simple (e.g., there is an order of difficulty on the problems, such as $1 < 3 < 2 < 4 < 5$), nor are we suggesting that the structure merely depends on the logical relationships between the problems. An example in another field may be helpful: If we discover that a student can name five of the nine judges of the Supreme Court of the United States, this student most likely also knows the name of the vice president of the United States, and that question need not be asked in a test evaluating the student's competency in the current U.S. political organization. Clearly, no logical relationship is involved in such an inference.

In general, we remark that, from the responses (correct or incorrect) to some problems in a given field, some likely inferences can be made regarding the responses to other problems in that field; these inferences could be used in the design of efficient knowledge assessment procedures. Obviously, we have computerized procedures in mind. In the last few years, our group has been engaged in developing a mathematical theory basic to this issue, together with the associated computer algorithms (Degreef, Doignon, Ducamp, & Falmagne, 1986; Doignon & Falmagne, 1985, 1987, 1988; Falmagne, 1989a, 1989b; Falmagne & Doignon, 1988a, 1988b; Koppen & Doignon, in press; Villano, Falmagne, Johannesen, & Doignon, 1987). The purpose of this article is to give a comprehensive description of our progress and to discuss in detail an empirical application of a basic model.

The central concept in our work is that of a knowledge state, by which we mean a set containing all the problems that some individual is capable of solving. The collection of all such states is called the knowledge structure. Specific assumptions focus consideration on a special class of knowledge structures called knowledge spaces. No cognitive interpretation is attached to these concepts. However, they can be shown to be consistent with some standard features of psychometric theory, such as skills or abilities (see Knowledge Spaces and Skills section).

Before providing details, we point out that our work has the potential to be applied in cases superficially very different from the assessment of knowledge. For example, consider a student as a system; the task at hand is then to uncover the state of this system by an appropriate sequence of verifications. Under this guise, obvious analogies appear between the assessment of knowledge and other situations, such as those outlined in the following paragraphs.

1. *Failure analysis.* Consider a complex device, such as a telephone interchange (or a computer, or a nuclear plant). At some point in time, the device's behavior indicates a failure. The system's administrator (or a team of experts) will perform a sequence of tests to determine the particular malfunction responsible for the difficulty.

2. *Medical diagnosis.* A physician examines a patient. To determine the disease (if any), the physician will check which symptoms are present. As in the preceding example, a carefully designed sequence of verifications will take place. Thus, the system is the patient, and the state is his or her medical condition. (For an example of a computerized medical diagnostic system, see Shortliffe, 1976, and Shortliffe & Buchanan, 1975.)

3. *Pattern recognition.* A pattern-recognition device analyzes a visual display to detect one of many possible patterns, each of which is defined by a set of specified features. Consider a case in which the presence of features is checked sequentially, until a pattern can be identified with an acceptable risk of errors. In this example, the system is a visual display, and the pattern is its state. (For a first contact with the vast literature on pattern recognition, consult, for example, Duda & Hart, 1973, and Fu, 1974.)

Except for occasional remarks, we do not pursue these analogies further in this article. However, the reader may want to keep these cases in mind while appraising our theoretical constructions.

Basic Concepts

Initially, the questions we asked ourselves were as follows: How can we use previous responses given by a subject to some problems to identify the remaining possible knowledge states? How can we use that information to choose the next problem to give the subject? More fundamental, how can we structure the set of problems to permit such inferences, and how can the concept of a knowledge state be defined? Some precise notation will be useful.

We denote by Q the complete set of problems under consideration. In our example, this set contains Problems $1, \ldots, 5$. In practice, the set Q may be quite large and contain dozens, maybe hundreds of problems. As indicated, we assume that the responses are coded only as correct or incorrect (the possible limitations of this convention are addressed in the Discussion section). The first formalization that comes to mind is based on the idea that, from observing that a student is capable of solving a given problem, one can sometimes surmise that this student can also solve other problems. This suggests the introduction of a binary relation $\precsim$, with the following interpretation:

$q \precsim t$ if and only if from a correct response to problem t we can surmise a correct response to problem q.

We shorten this as q can be *surmised from* t. The relation $\precsim$ will be called the *surmise relation.* This relation is obviously reflexive, and it is reasonable to suppose that it is also transitive. In other words, it is a *quasi-order* on the set Q. (For the terminology on order relations, see, for example, Roberts, 1979.)

A plausible surmise relation $\precsim$ for our five problems is shown in Figure 1. The conventions of the figure (known as a *Hasse diagram* in graph theory) are as follows. When some problem q can be reached from some other problem t by descending lines, this means that q can be surmised from t; that is, $q \precsim t$. For instance, from a correct response to Problem 5, we can surmise a correct response to Problems 2, 1, and 3 (thus, $2 \precsim 5$, $1 \precsim 5$, and $3 \precsim 5$). However, from a correct response to Problem 2, we can only surmise a correct response to Problem 1 (there is no descending line linking Problems 2 and 3). A complete list of the possible knowledge states can be inferred from this representation. The student may know nothing at all, or may know just Problem 1, or just Problem 3, or maybe Problems 1 and 3, or Problems 1, 2, and 3 (etc.). Notice that there is no state of knowledge containing exactly Problems 1, 3, and 5 (from the knowledge of Problem 5, we can infer the knowledge of Problem 2). There are 10 knowledge states out of a possible $2^5 = 32$: $\varnothing$, $\{1\}$, $\{3\}$, $\{1, 3\}$, $\{1, 2\}$, $\{1, 2, 3\}$, $\{1, 2, 4\}$, $\{1, 2, 3, 5\}$,

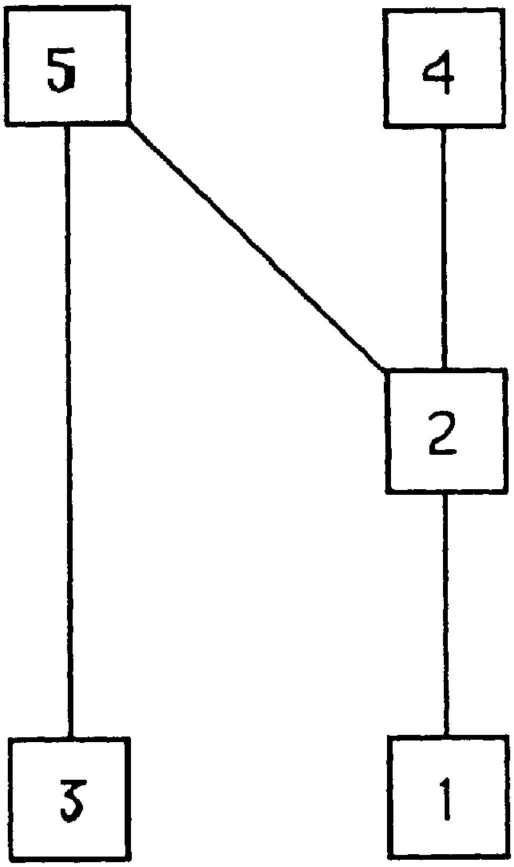

Figure 1. Hasse diagram of a plausible surmise relation of the five questions.

$\{1, 2, 3, 4\}$, and $\{1, 2, 3, 4, 5\}$. Formally, the definition of the knowledge state corresponding to a surmise relation $\precsim$ is

$$K \subseteq Q \text{ is a } state \Leftrightarrow (\text{for all } q, t \in Q,\ q \precsim t, \text{ and } t \in K \Rightarrow q \in K). \qquad (1)$$

In plain language: A set K of problems is called a *state* if whenever K contains some problem t, it also contains all the problems that can be surmised from t. There is an accidental feature of our example: There is no case in which two distinct problems can be surmised from each other. Formally, this means that the surmise relation $\precsim$ satisfies the *antisymmetry* condition $q \precsim t$ & $t \precsim q \Rightarrow q = t$. A quasi-order that is also antisymmetric is called a *partial order.* The antisymmetry condition is not required because it does not make sense as a general rule. We shall return to this point later in this section.

However appealing the simplicity of this formalization, it has one undesirable consequence: Any problem t has a unique set of prerequisites or antecedents, namely, the set of all problems that can be surmised from t. In the Hasse diagram, those problems can be reached from t by descending lines. For instance, the set of prerequisites of Problem 5 is the set $\{1, 2, 3\}$. This condition may be deemed acceptable in our example, but it is certainly absurd in general. Counterexamples are easy enough to manufacture, even in a highly structured field such as mathematics. Suppose that one discovers that a student is capable of solving a system of linear equations, for example, three equations in three unknowns. One may infer from this fact that the student is conversant with the concept of a determinant, or that the student is capable of inverting a matrix, or that the student knows how to manipulate equations (adding or subtracting them) as in the Gauss method. Moreover, any combination of these three hypotheses is possible. Whichever is the case is not determined.

Thus, the model is too strong. However, which assumption or assumptions should be dropped is by no means clear. At this juncture, a change of viewpoint will be profitable. What will be retained is the fundamental idea that the knowledge state of some individual is the set of all the problems that this individual is capable of solving (omitting, for the time being, complicating factors, such as incorrect responses due to careless errors and correct responses resulting from lucky guesses). The knowledge structure is then the collection of all the subsets of problems representing states. Thus, in our example, denoting by $\mathcal{F}$ the knowledge structure, we obtain

$$\mathcal{F} = \{\varnothing, \{1\}, \{3\}, \{1, 3\}, \{1, 2\}, \{1, 2, 3\}, \{1, 2, 4\},$$
$$\{1, 2, 3, 5\}, \{1, 2, 3, 4\}, \{1, 2, 3, 4, 5\}\} \quad (2).$$

Clearly, not all subsets of problems are suitable to represent states. To determine the kinds of constraints that may arise for a knowledge structure, one may ask: Is there anything special about the knowledge structure derived from a surmise relation that happens to be a quasi-order, such as the one represented in Figure 1? Notice that the collection of states specified by Equation 2 satisfies the following property:

If K and K' are states, then $K \cup K'$ is also a state.

For example, we have the two states $\{3\}$ and $\{1, 2\}$, but we also have the state $\{1, 2, 3\} = \{3\} \cup \{1, 2\}$. A family of states satisfying this property will be called *union closed,* or $\cup$-*closed.* Another property, in the same vein but involving intersections, requires that

if K and K' are states, then $K \cap K'$ is also a state.

If this property holds for some family of states, we say that it is *intersection closed,* or $\cap$-*closed.* From Equation 2, it is easy to verify that the family of states $\mathcal{F}$ of our example is both $\cup$-closed and $\cap$-closed. This example illustrates a general situation, which can be described in terms of the following classical result.

Theorem 1 (Birkhoff, 1937): For any set Q, the formula

$$q \precsim t \Leftrightarrow (t \in K \Rightarrow q \in K, \text{ for all } K \in \mathcal{F}) \qquad (3)$$

establishes a one-to-one correspondence between the set of all quasi-orders $\precsim$ on Q, and the collection of all families $\mathcal{F}$ of subsets of Q that are $\cup$-closed and $\cap$-closed.

(See also Monjardet, 1970, and Doignon & Falmagne, 1985.) In other words: For any surmise relation $\precsim$ on the set Q, which is reflexive and transitive, there is exactly one family of states $\mathcal{F}$ that is $\cup$-closed and $\cap$-closed, and vice versa. Moreover, the specifications of $\mathcal{F}$ by $\precsim$ and of $\precsim$ by $\mathcal{F}$ are given by Formula 3.

This result is useful in that it introduces a sharply different outlook on a collection of knowledge states. Previously, the defining condition was that a collection of states had to be consistent with (or specifiable in terms of) a surmise relation, which was assumed to be transitive and reflexive. Now, however, the surmise relation can be dispensed with. We can simply say that we are considering families $\mathcal{F}$ of knowledge states that are $\cup$-closed and $\cap$-closed. In this form, figuring out how the model can be weakened is much easier. The assumption that the collection $\mathcal{F}$ is $\cup$-closed seems, on the whole, quite reasonable. Consider a group of subjects engaged in extensive interactions over a period of time. Suppose that the knowledge states of Subjects $1, 2, \ldots, n$ are $K_1, K_2, \ldots, K_n$ (with respect to some body of information). At some point, one of the subjects conceivably will have accumulated all the knowledge contained in the group. That is, the knowledge state of that subject will be $K_1 \cup K_2 \cup \ldots \cup K_n$. We are not asserting that this will necessarily happen. However, having a state in the structure to cover this case certainly makes sense because this case might be realized. On the other hand, the assumption that the collection $\mathcal{F}$ is $\cap$-closed is much harder to justify and may be dropped whenever appropriate. In such a case, as we shall see in the Knowledge Spaces, Prerequisites, and Surmise Systems section of this article, any given problem may have several sets of prerequisites.

Another important idea must be introduced. Consider the case in which two particular problems, q and s, are contained in exactly the same states. This situation, which did not arise in our example of Equation 2, is illustrated by the collection of states $\mathcal{R} = \{\varnothing, \{q, s\}, \{q, s, t\}, \{t\}, \{t, r\}, \{q, s, t, r\}\}$. In the collection $\mathcal{R}$, problem q is in a given state if and only if problem s is in that state. If such a situation is observed with a large set of problems and a sufficiently rich collection of knowledge states, one tempting interpretation is that problems q and s are actually testing the same abstract notion. We shall go one step further and call *notion* any maximal subset of problems, all of which are contained in exactly the same states. Thus, the subset $\{q, s\}$ is a notion, the other notions being $\{t\}$ and $\{r\}$. At this stage, no cognitive theorizing is attached to this concept, the primary importance of which lies in the simplification of the data that it allows. Since q and s play exactly the same role vis-à-vis the states, to confound them makes sense. More generally, we lose nothing of importance by considering the notions, rather than the problems, as the basic entities. Consequently, we sometimes assume in the following sections that this simplification has been carried out.

Our discussion is summarized by the following fundamental definition, which is, unfortunately but unavoidably, somewhat abstract.

> *Definition 1:* A *knowledge structure* is a pair $(Q, \mathcal{K})$ in which Q is a set of problems or questions and $\mathcal{K}$ is a collection of subsets of Q, called the (*knowledge*) *states*.

Occasionally, when no ambiguity can arise regarding the set of problems, we refer to $\mathcal{K}$, rather than to $(Q, \mathcal{K})$, as the knowledge structure. The set of all states containing a particular problem q will be denoted by $\mathcal{K}_q$. The set of all problems belonging to the same states as q is denoted by $[q]$ and is called a notion. Thus,

$$[q] = \{r \mid \mathcal{K}_r = \mathcal{K}_q\}.$$

Clearly, the collection of all notions is a partition of the set of problems. As an illustration, we have, from Equation 2,

$$\mathcal{F}_4 = \{\{1, 2, 4\}, \{1, 2, 3, 4\}, \{1, 2, 3, 4, 5\}\}$$

and

$$[4] = \{4\},$$

because none of the other problems belong to exactly the same states as does Problem 4. A knowledge structure is said to be *discriminating* when every notion contains exactly one problem, that is, when $[q] = \{q\}$ for all problems q. (This is the case in the knowledge structure $\mathcal{F}$.)

A knowledge structure $(Q, \mathcal{K})$ is called a *knowledge space* when the following two axioms hold:

[K1] The set Q and the empty set $\varnothing$ are states.
[K2] Any union of states is a state (that is, $\mathcal{K}$ is $\cup$-closed).

Suppose that, in addition, the knowledge structure satisfies

[K3] Any intersection of states is a state (that is, $\mathcal{K}$ is $\cap$-closed).

Then, $\mathcal{K}$ will be called *quasi-ordinal*. A discriminating, quasi-ordinal space is a *partially ordinal* space. (It is easy to show that when a knowledge structure $\mathcal{K}$ is defined from a quasi-order $\lesssim$, then $\mathcal{K}$ is discriminating if and only if $\lesssim$ is a partial order, that is, $\lesssim$ is antisymmetric.)

The knowledge structure $\mathcal{F}$ is a partially ordinal space because Equation 2 defines a discriminating, $\cup$-closed, $\cap$-closed knowledge structure.

In the rest of this article, we consider for our five problems, as a working hypothesis, the knowledge space

$$\begin{aligned}\mathcal{K} = \{&\varnothing, \{1\}, \{3\}, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\},\\ &\{3, 4\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\},\\ &\{1, 2, 3, 4\}, \{1, 2, 3, 5\}, \{1, 2, 3, 4, 5\}\}, \end{aligned} \tag{4}$$

which does not satisfy the intersection property of Axiom [K3]. How such a knowledge space can be achieved will be explained in the Building the Knowledge Space section, in which a general procedure will be described. According to Birkhoff's theorem, this collection of states cannot be defined from a transitive, reflexive surmise relation because [K3] does not hold. Indeed, the set $\{2\} = \{1, 2, 3\} \cap \{1, 2, 4\} \cap \{2, 3, 4\}$ is not a state of $\mathcal{K}$. Excluding $\{2\}$ as a state is sensible: Any student capable of solving Problem 2 should be expected to have mastered either Problem 1, or Problem 3, or possibly both. A graphic representation of the knowledge space defined by Equation 4 is given in Figure 2.

This figure makes clear that the knowledge space $\mathcal{K}$ also satisfies a condition not encountered before. Namely, the lines of Figure 2 are always joining two "consecutive" states: The state to the right contains exactly one more problem than does the state to the left. This condition is not required for general knowledge spaces, but it plays a critical role in the section titled A Stochastic Learning Theory for a Knowledge Space in connection with the concept of a "learning path."

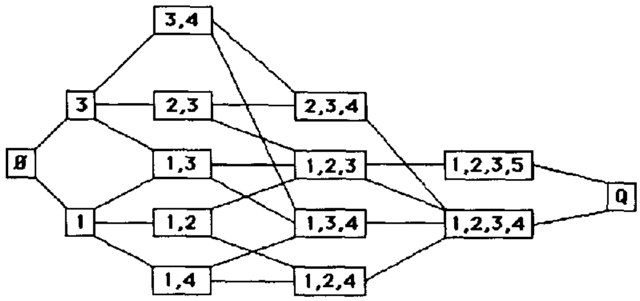

Figure 2. Graph of the knowledge space of Equation 4. (Each state is represented by a box, and a line joining two states, K and K', from left to right indicates that K' is the smallest state strictly including K.)

One might ask why so much attention is given to restrictive assumptions. Why not have as broad a theory as possible and simply deal with knowledge structures in general? The closure under union, the key axiom of a knowledge space, is appealing for several reasons. We have argued that this condition seemed realistic in some empirical situations. Moreover, the concept of a knowledge space appears to be consistent with a more traditional model formalized in terms of skills (see the Knowledge Spaces and Skills section). It is also closely related to the AND/OR graphs of artificial intelligence (see Doignon & Falmagne, 1985). Some practical considerations are relevant. The concept of a knowledge space can be shown to be equivalent to another one, which allows a very convenient method of acquiring information from experts. This means that the actual construction of the knowledge space, at least as a working hypothesis, is feasible (see the Building the Knowledge Space section). Finally, when a knowledge structure is a knowledge space, an efficient way of storing (in a computer's memory) the list of knowledge states is available. We illustrate this fact with our standard example. Notice that all 15 states given in Equation 4 can be obtained by taking all arbitrary unions (including the empty union) of the states included in the subcollection

$$\mathcal{B} = \{\{1\}, \{3\}, \{1, 2\}, \{1, 4\}, \{2, 3\}, \{3, 4\}, \{1, 2, 3, 5\}\}, \quad (5)$$

which only contains 7 states. Such a reduction can always be achieved. That is, when the set Q of problems is finite, it is always possible to find a "minimal" subcollection of states $\mathcal{B}$ such that all the states in the knowledge space $\mathcal{K}$ can be found by taking unions of states in $\mathcal{B}$. The subcollection $\mathcal{B}$ is called *minimal* in the sense that it is included in any other subcollection of states generating the states in $\mathcal{K}$ by taking all possible unions. There is always exactly one subcollection $\mathcal{B}$ satisfying these properties; we call it the *basis* of $\mathcal{K}$.

Nevertheless, the $\cup$-closure assumption may not be fulfilled in some practical applications. Even in such a case, the main lines of the general theory developed herein would stand. A full discussion of this issue is postponed until the Discussion section.

This approach to the representation of knowledge, as a step toward assessing the knowledge states of individuals, prompts several basic issues, which are discussed in the various sections of this article.

We started our discussion with a model in which the possible knowledge states were derived from a binary surmise relation $\precsim$ on the set Q of problems, which was assumed to be reflexive and transitive. This model had the undesirable feature that to each problem there corresponded a unique set of prerequisites, and the model was rejected for that reason. The concept of a knowledge space is more general and escapes this criticism; this will be shown in the next section, where the exact relationship between the collection of states and the prerequisites is described.

There is a long-standing tradition in psychology to analyze the results of mental tests in terms of psychometric models, the core of which is a representation (uni- or multidimensional, often numerical) of the concept of ability (Lord, 1974; Lord & Novick, 1974; Wainer & Messick, 1983; Weiss, 1983). In principle, knowledge clearly can be assessed in terms of such models, and this approach may be sensible when one is interested in broad, long-term predictions of an individual's achievements. However, when the aim of the assessment procedure is an accurate description of the current knowledge of a specific body of information, the concept of ability seems a priori to be a costly detour. In any event, the choice of the most successful model ultimately will be a matter of empirical comparison. Such a comparison is an item in our research program, but it is not undertaken in this article. Beyond these utilitarian considerations, one can also ask whether our approach is conceptually consistent with more traditional ones on the basis of the concepts of skills or abilities. Some possible relationships are explored in the Knowledge Spaces and Skills section.

How can a knowledge space be constructed in practice? In our view, a first sketch of the space must be obtained by systematically consulting experienced teachers. However, we cannot simply ask these experts to give us a list of the possible knowledge states. For one reason, the concept of a knowledge state is somewhat abstract and may be difficult to convey exactly. For another, such a list will certainly contain, in any realistic application, thousands of states. Thus, what are the right questions to ask the experts? In the Building the Knowledge Space section, we show that the concept of a knowledge space is actually equivalent to another concept in which the knowledge states do not appear explicitly: This second concept is much more suitable for acquiring the relevant information from experts. We outline a practical procedure, which we illustrate by an application involving our five problems and the knowledge space represented in Figure 2.

This first sketch of a knowledge space, based on experts' opinions, need not be taken for the final word and should be confronted with experimental data. At this stage, the knowledge space is regarded as a model capable of making predictions concerning the actual responses of a sample of subjects to the problems. However, such data are typically "noisy"—for instance, a subject might make a careless error in a computation or might guess the correct response. If the knowledge space is essentially correct, the states will thus govern the subject's responses only through a mediation by some random process that remains to be specified. Moreover, the standard evaluations of the goodness of fit of a model in a noisy situation are based on statistical techniques. The correct approach is to build a probabilistic model, the core of which is the knowledge space, and to specify, up to the values of some parameters, the probabilities of all possible patterns of responses (there are $2^5 = 32$ such patterns in our example). One such model, which seems especially interesting, is described under the heading of A Stochastic Learning Theory for a Knowledge Space. The model is simple enough to allow explicit predictions and offers an appealing explanation of a complex set of data. The model has been applied to the responses to the five problems given by 497 high school students. This experiment and the results of the statistical analysis are described in detail in the section An Application of the Learning Theory. We show how such an analysis leads—via likelihood ratio techniques—to testing possible simplifications of the knowledge space.

In practice, how can we uncover the knowledge state of an individual, once an acceptable knowledge space has been constructed? Indeed, the ultimate goal of this work is the construction of practical, computerized knowledge assessment procedures. In the Uncovering the State of a Student section, two classes of such procedures are outlined within the framework of Markovian processes, and we show how such procedures can be used to determine, at least to a good approximation, the state of an individual.

The article ends with a discussion of some potential criticisms to our approach and an outline of further developments.

Knowledge Spaces, Prerequisites, and Surmise Systems

Intuitively, a set of prerequisites for some problem q is a minimal set of problems that must have been mastered before tackling Problem q. We claim that, in a knowledge space, any problem may have more than one set of prerequisites. Our discussion of this fact is based on our standard example and, more specifically, on the knowledge space obtained in the previous section, namely

$$\begin{aligned}\mathcal{K} = \{&\varnothing, \{1\}, \{3\}, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\},\\ &\{3, 4\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\},\\ &\{1, 2, 3, 4\}, \{1, 2, 3, 5\}, \{1, 2, 3, 4, 5\}\}\end{aligned}$$

with basis

$$\mathcal{B} = \{\{1\}, \{3\}, \{1, 2\}, \{1, 4\}, \{2, 3\}, \{3, 4\}, \{1, 2, 3, 5\}\}.$$

Suppose that all we know about a particular student, named John, is that he has solved Problem 4. (We still assume that no careless errors or correct guessing are involved. This assumption is dropped in the Stochastic Learning Theory section.) What can we infer from this observation regarding John's knowledge state? Surely, his state must contain Problem 4. Thus, this state must be a member of

$$\begin{aligned}\mathcal{K}_4 = \{&\{1, 4\}, \{3, 4\}, \{1, 2, 4\}, \{1, 3, 4\},\\ &\{2, 3, 4\}, \{1, 2, 3, 4\}, \{1, 2, 3, 4, 5\}\}.\end{aligned}$$

However, we cannot infer that John's state contains Problem 2 or Problem 5: We have no evidence for such conclusions. The only reasonable inference that can be made is that John's state contains Problem 1 or Problem 3: Any state containing Problem 4 contains Problem 1 or Problem 3. Thus, we have two possible sets of prerequisites for Problem 4, namely $\{1\}$ and $\{3\}$. In other terms, John's state includes $\{1, 4\}$, or $\{3, 4\}$, or possibly both sets. These two sets will be called the *clauses for* Problem 4. Notice that these two clauses are knowledge states. They are actually the "minimal" states containing Problem 4 (to say that a state K is a *minimal* state containing a problem q means that K contains q and that there is no other state $K' \subset K$ containing q).

Table 1
Clauses for Each of the Five Problems

Problem	Clause
1	$\{1\}$
2	$\{1, 2\}, \{2, 3\}$
3	$\{3\}$
4	$\{1, 4\}, \{3, 4\}$
5	$\{1, 2, 3, 5\}$

In the sequel, we focus on the clauses rather than on the set of prerequisites. (The possible sets of prerequisites for some problem q are automatically obtained by removing q from the clauses for q.) In general, if a subject is found to have mastered a given problem, then at least one of the clauses for the problem must be included in the subject's state. The clauses for each of the five problems are listed in Table 1. Notice that each of the clauses is an element of the basis $\mathcal{B}$. Conversely, each element of the basis is a clause for some problem. This example illustrates a general situation. Let $\tilde{\mathcal{K}}_q$ represent the set of all minimal states containing some problem q in a knowledge space $\mathcal{K}$ with basis $\mathcal{B}$. Thus, $\tilde{\mathcal{K}}_q \subseteq \mathcal{K}_q$, the set of all states containing q. We also have $\tilde{\mathcal{K}}_q \subseteq \mathcal{B}$ for any problem q, and the set $\sigma(q)$ of all the clauses for q is obtained by setting

$$\sigma(q) = \tilde{\mathcal{K}}_q. \tag{6}$$

As another illustration, consider $\sigma(2) = \{\{1, 2\}, \{2, 3\}\} = \tilde{\mathcal{K}}_2$.

Equation 6 specifies how the clauses can be constructed from the basis of a knowledge space. Notice that such clauses automatically satisfy three conditions. We have already encountered one of them, namely

[S1] Every clause for a problem q contains q.

The next condition seems a natural requirement for the concept of a clause as we understand it: A clause for a problem q is a minimal set of problems that has to be mastered before problem q can be solved. However, suppose that C is such a clause for q and that there is another problem q' contained in C. Surely, then, q' is on the path to the mastery of q. This means that if, for example, problem q' has three clauses, at least one of them must be included in C (Figure 3). In general, we have the following condition, which also results from Equation 6.

[S2] If a problem q' is in some clause C for q, then there must be some clause C' for q' included in C.

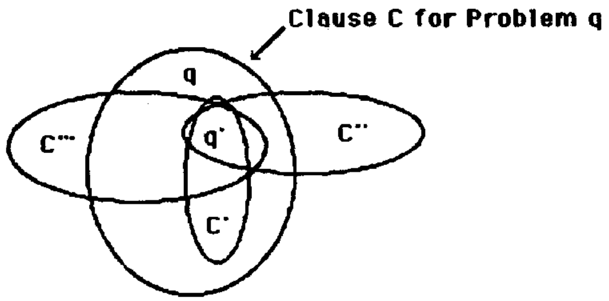

Figure 3. Illustration of Condition [2]. (Problem q' is in clause C of problem q and has the three clauses C', C'' and C'''. At least one of these three clauses must be included in C. Shown is $C' \subseteq C$.)

Notice that we may have $C = C'$; thus, C is a clause for both q and q'. An example is provided by the knowledge space

$$\mathcal{D} = \{\varnothing, \{a, b\}, \{b, c\}, \{a, c\}, \{a, b, c\}\},$$

with basis

$$\{\{a, b\}, \{b, c\}, \{a, c\}\}.$$

The clauses for a are $\{a, b\}$ and $\{a, c\}$. We have $b \in \{a, b\}$, and $\{a, b\}$ is also a clause for b. Condition [S2] is satisfied.

The third condition eliminates a situation in which we would have two clauses, C and C', for the same problem q, with C' a proper subset of C. From observing a correct response to problem q, we have no reason to take $C - \{q\}$ as a possible set of prerequisites, because in our interpretation of the clauses, the mastery of all the problems in the smaller set $C' - \{q\}$ is sufficient to access problem q. The next condition states that such cases do not arise.

[S3] Any clause C for q is minimal in the sense that if C' is also a clause for q and $C' \subseteq C$, then $C' = C$.

Any set of clauses defined from a knowledge space through Equation 6 satisfies these three conditions.

Could we go the other way around? That is, suppose that, for each problem q, we had the collection $\sigma(q)$ of all of its clauses and that these clauses satisfy Conditions [S1], [S2], and [S3], now considered as axioms. Could we, then, reconstruct the knowledge space exactly? The answer is yes. The reader may notice the resemblance of this situation with that encountered in the last section in connection with Birkhoff's result (Theorem 1). On the one hand, there were families of states that were $\cup$-closed and $\cap$-closed, and, on the other hand, there were surmise relations $\precsim$ that were quasi-orders, that is, which were reflexive and transitive. Birkhoff's theorem stated, then, that these two concepts were equivalent. The situation discussed in this section is similar, but it concerns an equivalence between knowledge spaces and systems of clauses satisfying the three Axioms [S1], [S2], and [S3]. The next definition summarizes the discussion presented thus far and paves the way for a precise statement of that equivalence in Theorem 2.

Definition 2: Let Q be a set of problems, and let σ be a function that associates to each element q of Q a nonempty collection $\sigma(q)$ of subsets of Q called the clauses for q. Suppose that these clauses satisfy Axioms [S1], [S2], and [S3]. Then, the pair (Q, σ) is called a *surmise system,*[1] and σ is a *surmise function on* Q.

Theorem 2: Suppose that Q is a finite set of problems. Then there is a one-to-one correspondence between the set of all surmise functions σ on Q and the set of all knowledge spaces $\mathcal{H}$ on Q. Moreover, this correspondence is specified by Equation 6.

This result, which generalizes Birkhoff's theorem in the finite case, is due to Doignon and Falmagne (1985). A proof can be found in their article.

Knowledge Spaces and Skills

So far, our approach to the representation of knowledge has been deliberately free of any cognitive interpretation. Our main concepts do not rely on traditional explanatory features of psychometric research, such as "latent traits" in the guise of skills or abilities. However, the concepts can be reconciled with such features. We briefly explore some of the possible relationships. Following Marshall (1981), for example, we assume the existence of some abstract set S of skills. For each problem $q \in Q$, there is a corresponding subset $f(q) \subseteq S$ of skills, namely those very skills required for (or at least relevant in) solving the problem. Thus, f is a function mapping Q into the set of all subsets of S. (Marshall draws an important distinction between component skills and their combinations. Such a distinction is valid if the set of skills is specified. Otherwise, one can always consider any combination of skills as just another skill. This is the position we adopted herein.) Implicitly, the function f defines a hierarchy between the items: If all the skills required to solve problem q are also required to solve problem q', then, from observing that a subject is capable of solving problem q', one should surmise that this subject is also capable of solving problem q. This suggests the definition of a binary relation $\mathbf{P}$ on Q by the equivalence

$$q \, \mathbf{P} \, q' \Leftrightarrow f(q) \subseteq f(q').$$

Obviously, such a relation is a quasi-order closely related to the surmise relation $\precsim$ discussed in the first section. (In principle, only the method of construction is different.) As in Equation 1, the relation $\mathbf{P}$ can be used to generate the knowledge states. However, the relation $\mathbf{P}$ then suffers from the same defects as the surmise relation $\precsim$ in that it implies the existence of a unique set of prerequisites for each problem. This suggests making the weaker assumption that several sets of skills may be attached to each problem, every one of which may be used to solve the problem. In this framework, if we discover that some individual is capable of solving a given problem, we cannot automatically infer which skills this individual possesses, because there may be more than one way to arrive at a solution. The concept of a latent trait is germane to a distinction made in linguistic theory between competence and performance (Chomsky, 1965). We define *competency* to be a complete set of skills sufficient to

[1] Our terminology in this definition differs slightly from that used in Doignon and Falmagne (1985), in which a surmise system was labeled a *space-like surmise system.*

Table 2
Hypothetical Competencies Associated With Each of the Five Problems

Problem	Competency
1	{**a**, **b**}
2	{**a**, **b**, **d**}, {**c**, **d**}
3	{**c**}
4	{**a**, **b**, **e**}, {**c**, **e**}
5	{**a**, **b**, **c**, **d**, **f**}

solve a problem. Thus, to any problem there is associated a collection of competencies. The key step consists in specifying from the full collection of competencies the subsets of problems forming knowledge states; thus, the following definition seems sensible:

> *Definition 3:* A subset K of problems is a knowledge state if and only if there is a subset M of skills such that K contains all those problems having at least one competency included in M and only those problems. We shall then say that K and the elements of K are *generated by M*.

Some hypothetical competencies associated with each of the five problems are shown in Table 2. We assume that six skills are involved: $S = \{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{e}, \mathbf{f}\}$.

It can easily be checked that the 15 states of the knowledge space $\mathcal{K}$ of Equation 4 are generated by applying Definition 3 to the competencies listed in Table 2. For instance, $\varnothing \subseteq Q$ is a state because it is generated by $\varnothing \subseteq S$; $\{1, 2\}$ is a state because it is generated by $\{\mathbf{a}, \mathbf{b}, \mathbf{d}\}$; $\{2, 3, 4\}$ is a state because it is generated by $\{\mathbf{c}, \mathbf{d}, \mathbf{e}\}$; and so forth. On the other hand, the 17 nonstates can be obtained from three facts:

1. Any subset of skills generating some state K containing 2 includes $\{\mathbf{a}, \mathbf{b}, \mathbf{d}\}$ or $\{\mathbf{c}, \mathbf{d}\}$. Thus, K also contains 1 or 3.
2. Any subset of skills generating some state K containing 4 includes $\{\mathbf{a}, \mathbf{b}, \mathbf{e}\}$ or $\{\mathbf{c}, \mathbf{e}\}$. Thus, K also contains 1 or 3.
3. Any subset of skills generating some state K containing 5 includes $\{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{f}\}$. Thus, K also contains 1, 2 and 3.

Notice that in Table 2, two skills are in some sense redundant: Skills **a** and **b** belong to exactly the same competencies. We could, without loss of predictive power, decide to combine these two skills into one.

In this particular case, we have assumed that the set of competencies was finite, but this is by no means necessary. For instance, a quantitative form of a competency might be the set of all numbers $t \leq \theta$, with θ representing some ability level.

The point of this simple example is to show that an interpretation of a knowledge space in terms of a structure of underlying skills is sometimes possible. (It turns out that such a construction is always feasible. We shall not prove this fact herein, however.) The theory of this type of interpretation remains to be developed. The types of questions that this theory would address are: Under which conditions on the competencies do the states (as specified by Definition 3) form a knowledge space? What is the exact structural relation between the knowledge space and the competencies? The collection of competencies displayed in Table 2 bears a striking resemblance to the clauses given in Table 1 (if the redundant skills **a** and **b** are combined, the left column in Table 2 becomes just a relabeling of the corresponding column in Table 1). Is this resemblance fortuitous, or does it illustrate an essential relationship? These questions will be discussed in a later publication if the results are of interest.

Building the Knowledge Space

Even for sets of moderate size (e.g., Q contains 30 problems), the practical construction of the knowledge space is a considerable enterprise. This construction must rely partly on experimentation, which means collecting the frequencies of occurrence of all the possible patterns of responses in a large sample of subjects and, through some appropriate analysis, inferring the knowledge space that can best account for the data. An example of what we have in mind will be given in the next two sections of this article. In our view, this is only the second step to be taken. There is much to be gained by first obtaining a sketch of the knowledge space based on expert teachers' opinions. However, as already argued, asking a teacher, no matter how experienced, whether particular sets of problems are possible knowledge states does not seem feasible. An indirect approach must be taken.

There is, however, a type of question that an expert will fully comprehend and might also be able to answer reliably:

> [Q1] Suppose that a student under examination has just provided a wrong response to problem q_1. Is it practically certain that this student will then also fail problem q_2? We assume that careless errors and lucky guesses are excluded.

For expository purposes, we simplify this to the question: Does failing q_1 entail failing q_2?

The teacher can draw from his or her educational experience with a population of students to decide whether to respond "yes" or "no." A positive response would mean that some sets of problems can be eliminated as possible states, namely all those sets containing problem q_2 but not problem q_1. Indeed, if there are some states in the knowledge space containing q_2 and not q_1, then there might be a student failing q_1 but not q_2, and the expert should respond "no" to [Q1]. A positive response to [Q1] will be coded as $q_1 \mathrm{P} q_2$, and a negative response will be coded as $q_1 \mathrm{N} q_2$. Notice that not all such questions need to be asked. For instance, if we have observed $q_1 \mathrm{P} q_2$ and $q_2 \mathrm{P} q_3$, we can infer $q_1 \mathrm{P} q_3$ (if failing q_1 entails failing q_2 and failing q_2 entails failing q_3, then, surely, failing q_1 entails failing q_3). Thus, there is no need to ask [Q1] for the pair (q_1, q_3). This means that P, regarded as a binary relation, is transitive on the set Q of problems. This transitivity property can be used to considerably facilitate the task of the expert. Unfortunately, if the only questions asked were of Type [Q1], we would only be able to construct the knowledge space in some very special cases, namely, cases in which the space is both ∪-closed and ∩-closed (this follows from Birkhoff's theorem). Clearly, some information might be missing. Consider Problems 1, 2, and 3 in our standard example, and suppose that the following two questions were asked: Does failing 1 entail failing 2? Does failing 3 entail failing 2? Assume that the expert is implicitly relying on the knowledge

Table 3
First 20 Steps of the Procedure for Questioning the Expert, in the Case of Knowledge Space $\mathcal{K}$ of Equation 4

Step t	$\lvert N_t \rvert$	$\lvert P_t \rvert$	Observed response	Pairs added to N_t	Pairs added to P_t	Sets deleted from $\mathcal{K}'$
0	0	80	—	—	—	—
1	1	80	1N 2	$1 \mid 2$	—	—
2	2	80	1N 3	$1 \mid 3$	—	—
3	3	80	1N 4	$1 \mid 4$	—	—
4	9	88	1P 5	$5 \mid 2, 5 \mid 3, 5 \mid 4, 1\,5 \mid 2, 1\,5 \mid 3, 1\,5 \mid 4$	$1 \mid 5, 1\,2 \mid 5, 1\,3 \mid 5, 1\,4 \mid 5, 1\,2\,3 \mid 5, 1\,2\,4 \mid 5, 1\,3\,4 \mid 5, 1\,2\,3\,4 \mid 5$	$\{5\}, \{2, 5\}, \{3, 5\}, \{4, 5\}, \{2, 3, 5\}, \{2, 4, 5\}, \{3, 4, 5\}, \{2, 3, 4, 5\}$
5	10	88	2N 1	$2 \mid 1$	—	—
6	11	88	2N 3	$2 \mid 3$	—	—
7	12	88	2N 4	$2 \mid 4$	—	—
8	16	92	2P 5	$5 \mid 1, 2\,5 \mid 1, 2\,5 \mid 3, 2\,5 \mid 4$	$2 \mid 5, 2\,3 \mid 5, 2\,4 \mid 5, 2\,3\,4 \mid 5$	$\{1, 5\}, \{1, 3, 5\}, \{1, 4, 5\}, \{1, 3, 4, 5\}$
9	17	92	3N 1	$3 \mid 1$	—	—
10	18	92	3N 2	$3 \mid 2$	—	—
11	19	92	3N 4	$3 \mid 4$	—	—
12	22	94	3P 5	$3\,5 \mid 1, 3\,5 \mid 2, 3\,5 \mid 4$	$3 \mid 5, 3\,4 \mid 5$	$\{1, 2, 4, 5\}, \{1, 2, 5\}$
13	23	94	4N 1	$4 \mid 1$	—	—
14	24	94	4N 2	$4 \mid 2$	—	—
15	25	94	4N 3	$4 \mid 3$	—	—
16	26	94	4N 5	$4 \mid 5$	—	—
17	27	94	12N 3	$12 \mid 3$	—	—
18	28	94	12N 4	$12 \mid 4$	—	—
19	28	97	13P 2	—	$1\,3 \mid 2, 1\,3\,4 \mid 2, 1\,3\,5 \mid 2$	$\{2\}, \{2, 4\}$
20	28	100	13P 4	—	$1\,3 \mid 4, 1\,3\,5 \mid 4, 1\,2\,3 \mid 4$	$\{4\}$

Note. The first column indicates the step number; the second column contains the current number of negative responses and inferences; the third column contains the current number of positive responses and inferences; the fourth column lists the observed response at the current step number; the fifth column contains the negative inferences drawn from the observed response (including the response itself, if it was negative); the sixth column contains the positive inferences drawn from the observed response (including the response itself); and the seventh column shows which states are deleted from the current knowledge space if a positive response is obtained (at the start—Step 0—*all* subsets of problems are considered states). Dashes indicate nothing is added or deleted.

space $\mathcal{K}$ of our example, which is specified in Equation 4. Since there are states in $\mathcal{K}$ containing 2 and not 1, and (other) states containing 2 and not 3, the expert should respond "no" to both questions. However, there are no states in $\mathcal{K}$ containing 2 but neither 1 nor 3, so the expert would actually respond "yes" to the question: Does failing both 1 and 3 entail failing 2? This positive response would eliminate $\{2\}$, $\{2, 4\}$, $\{2, 5\}$, and $\{2, 4, 5\}$ as possible states. It will be shown that if only questions of Type Q1 were asked, we would end up with the following collection of states:

$$\mathcal{K}' = \{\varnothing, \{1\}, \{2\}, \{3\}, \{4\}, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\}, \{1, 2, 3, 4\}, \{1, 2, 3, 5\}, \{1, 2, 3, 4, 5\}\}.$$

(This can be seen in Table 3. The sets $\{2, 5\}$ and $\{2, 4, 5\}$ have been removed on line 4 of the table, from the Observation 1P5.) Notice that $\mathcal{K}'$ is both $\cup$-closed and $\cap$-closed and contains all the right states; however, it also contains the three "useless" states $\{2\}$, $\{4\}$, and $\{2, 4\}$.

This discussion suggests a more general type of question for acquiring information from the expert:

[Q2] Suppose that a student under examination has just provided wrong responses to all the problems in a subset A of Q. Is it practically certain that this student will then also fail problem q in Q? We assume that careless errors and lucky guesses are excluded.

Extending our notation, we shall code positive and negative responses to such questions as $A \mathrm{P} q$ and $A \mathrm{N} q$, respectively. A positive response to the question: Does failing both 1 and 3 entail failing 2? will thus be denoted by $\{1, 3\}$P2. It can be established mathematically that, if all the questions of Type [Q2] are asked, then the knowledge space can be recovered exactly. This result, which is not proved herein (see Koppen & Doignon, in press), can be stated in the form of an equivalence between knowledge spaces and relations P, which is conceptually germane to the equivalences obtained in Theorems 1 and 2. In the rest of this section, we illustrate these ideas and results by progressively constructing the knowledge space $\mathcal{K}$ of Equation 4 from an appropriate sequence of questions of Type [Q2].

Our first step consists in ordering the questions to be asked of the experts; that is, we have to order the pairs (A, q) corresponding to the questions: Does failing all the problems in A entail failing problem q? It makes sense to begin with the simplest questions. We order the questions by the size of the set A in the pair (A, q). Thus, we ask (A, q) before (A', q') whenever $|A| < |A'|$. (Herein, and in the sequel, we denote by $|S|$ the number of elements in a set S.) This means that all the questions of Type [Q1] will be asked in the first stage of the procedure.

For sets A of the same size, a different kind of rule will be used, based on the fact that numbers have been arbitrarily assigned to the problems. (The labels q and q' denoting the questions are positive integers.) Suppose that $|A| = |A'| = n$: for example,

$$A = \{q_1, \ldots, q_n\}, \quad A' = \{q'_1, \ldots, q'_n\}.$$

Then,

$$(A, q) \text{ is asked before } (A', q')$$

if

$$\begin{aligned} \text{either } q_1 < q'_1 \quad &\text{or} \quad (q_1 = q'_1 \text{ and } q_2 < q'_2) \\ &\text{or} \quad (q_1 = q'_1 \text{ and } q_2 = q'_2 \text{ and } q_3 < q'_3) \\ &\text{or} \quad \ldots \\ &\text{or} \quad (q_1 = q'_1, \ldots, q_n = q'_n \text{ and } q < q'). \end{aligned}$$

(The type of ordering defined in the above formula is usually referred to as *lexicographic* by analogy with the order of the words in a dictionary.) Thus, for example, $(\{3, 5\}, 1)$ is asked before $(\{1, 2, 3\}, 5)$ because $|\{3, 5\}| = 2 < |\{1, 2, 3\}| = 3$, and $(\{2, 3\}, 4)$ is asked before $(\{3, 4\}, 1)$ because $|\{2, 3\}| = 2 = |\{3, 4\}|$, and $2 < 3$. Counting the maximal number of questions that might be needed is useful in designing a stopping rule. In principle, a question (A, q) may be constructed from any nonempty subset A of Q and any problem q. Writing m for the number of problems in Q, we thus have a maximum of $(2^m - 1) \times m$ questions; that is, $(2^5 - 1) \times 5 = 155$ questions in our example. However, many of these questions have positive responses known in advance. For example, Question $(\{1, 2, 3\}, 3)$ must yield a yes response because it means Does failing all three Problems 1, 2, and 3 entail failing Problem 3? In general, we must have

$$q \in A \Rightarrow A\mathrm{P}q. \tag{7}$$

The count of these positive responses known a priori is easy to make because it only involves counting the elements in the membership relation $\in$, for all the nonempty subsets of Q, or, equivalently, for all the subsets of Q. Since each subset of Q, together with its complement, contributes exactly m to this number, we obtain $2^m \times m/2 = m \times 2^{m-1}$, which becomes $5 \times 2^4 = 80$ in our example. We denote by P_0 the set of all questions with positive responses known a priori; thus, P_0 is just the (inverse relation of the) set membership relation $\in$ appearing to the left of the implication sign "$\Rightarrow$" in Formula 7. By symmetry, we denote by N_0 the set of all questions with negative responses known a priori. Because there are no such questions, we have

$$|P_0| = m \times 2^{m-1}, \quad |N_0| = 0.$$

In our example, we obtain $|\mathrm{P}_0| + |\mathrm{N}_0| = 80 + 0 = 80$.

Our method for questioning the expert proceeds by successively adding to P_0 or N_0 the positive or negative responses observed or inferred, yielding the two sequences of relations

$$\mathrm{P}_0 \subseteq \mathrm{P}_1 \subseteq \ldots \subseteq \mathrm{P}_t \subseteq \ldots,$$
$$\mathrm{N}_0 \subseteq \mathrm{N}_1 \subseteq \ldots \subseteq \mathrm{N}_t \subseteq \ldots.$$

The procedure stops when there are no more undecided questions, that is, when for some step number s

$$|\mathrm{P}_s| + |\mathrm{N}_s| = (2^m - 1) \times m.$$

We illustrate this procedure using our standard example. We assume that the expert's responses are consistent with the knowledge space of Equation 4. The first 20 steps of the procedure are represented in Table 3. In that table and in the sequel, we abbreviate $\{q_1, \ldots, q_n\}\mathrm{P}q$ as $q_1 \ldots q_n\mathrm{P}q$, with a similar convention for the relation N. We also refer to a question $(\{q_1, \ldots, q_n\}, q)$ by the shorthand notation $q_1 \ldots q_n|q$. The first column of the table contains the step number. The second and third columns contain the current values of $|\mathrm{N}_t|$ and $|\mathrm{P}_t|$; thus, for Step 0, the entries are 0 and 80, respectively. The fourth column lists the observed response from the expert, which will be of the form $q_1 \ldots q_n\mathrm{P}q_{n+1}$ or $q_1 \ldots q_n\mathrm{N}q_{n+1}$. The next two columns contain the pairs being added to the relations N_t and P_t. The symbol $\mathcal{K}^t$ in the heading of the last column denotes the collection of sets remaining at step t. The entry in this last column lists the sets being deleted from $\mathcal{K}^t$ at Step t, yielding $\mathcal{K}^{t+1}$. Note that $\mathcal{K}^0$ is just the collection of all subsets of Q. (At the start of the procedure, all subsets are potential states.)

We begin by asking Question $1|2$. Because there are states containing Problem 2 but not Problem 1, the response is negative. Therefore, 1N2 is recorded in the fourth column of Line 1 in the table, and the pair $1|2$ is added to N_0, forming the relation $\mathrm{N}_1 = \{1|2\}$. Because no positive response is observed or inferred and no sets are deleted from $\mathcal{K}^0$, we set $\mathrm{P}_1 = \mathrm{P}_0$ and $\mathcal{K}^1 = \mathcal{K}^0$. Next, we consider Questions $1|3$ and $1|4$. Again, two negative responses must be obtained. At the end of the third step of the procedure, we have

$$|\mathrm{N}_3| = 3, \ |\mathrm{P}_3| = 80, \ \mathrm{N}_3 = \{1|2, 1|3, 1|4\},$$
$$\mathrm{P}_3 = \mathrm{P}_0, \text{ and } \mathcal{K}_3 = \mathcal{K}^0.$$

The first positive response is observed at Step 4, with Question $1|5$. Consistent with the fact that the knowledge space has no state containing 5 but not 1, the expert gives a yes response. We add $1|5$ to P_4 in column 6. However, we can also add to P_4 all the pairs of the form $1q_1 \ldots q_n|5$, because failing all the problems in the set $\{1, q_1, \ldots, q_n\}$ certainly involves failing Problem 1. This justifies the additions of $1\,2\,|5$, $1\,3|5, \ldots, 1\,2\,3\,4|5$. There are many other consequences to be derived from this positive response. In particular, several a priori possible states may be discarded, that is, all states containing 5 but not 1: $\{5\}$, $\{2, 5\}$, $\{3, 5\}$, $\{4, 5\}$, $\{2, 3, 5\}$, $\{2, 4, 5\}$, $\{3, 4, 5\}$, and $\{2, 3, 4, 5\}$.

Thus, eight potential states were eliminated from the 32 initial ones, as indicated on Line 4 of Table 3 in the last column. Finally, we consider the addition of pairs to N_4 in the fifth column. Notice that there is no need to ask Questions $5|2$, $5|3$, and $5|4$ because from 1P5, together with the formulas 1N2, 1N3, and 1N4 obtained in the first three lines of the table, we can logically infer 5N2, 5N3, and 5N4.

The argument proceeds by contradiction. Assume, for example, that 5N2 does not hold. This means that a positive response must be obtained for the pair $5|2$, which is written as 5P2. Consider the two facts 1P5 and 5P2. By the transitivity of P on the set Q, we derive 1P2, contradicting the negative response 1N2 obtained in Line 1 of the table. The same reasoning applies to the two other pairs, 5|3 and 5|4. Thus, the three Pairs 5|2, 5|3, and 5|4 are added to N_3. However, Pairs 1 5|2, 1 5|3, and 1 5|4 can also be added. The argument is similar. Suppose, for example, 15P2. We have observed 1P5, from which we can deduce that failing 1 entails failing 1 and 5. However, failing 1 and 5, we have assumed, entails failing 2. Thus, failing 1 entails failing 2; that is, once again 1P2, contradicting our previous result 1N2. We see that, from the single positive response, 1P5, we managed to delete eight potential states, and 13 possible questions.

We shall not analyze Table 3 any further. The discussion of these few examples should be sufficient for the reader to understand the basic mechanisms of the procedure.

The knowledge space is obtained at the 20th question. However, the experimenter will know this fact only after the 29th question, when noticing that

$$|N_{29}| + |P_{29}| = 53 + 102 = 155 = 31 \times 5.$$

After Step 20, all further questions receive a no response.

One may ask whether applying such a procedure in a practical situation, involving many problems, is realistic. One could easily construct examples requiring a considerable number of questions to obtain the knowledge structure. However, the collection of states $\mathcal{H}^t$ remaining at Step t of the procedure (for any value of t) is a knowledge space and contains all the states belonging to $\mathcal{H}$, the knowledge space to be found. The procedure can thus be stopped at any time, whenever the cost becomes prohibitive. In particular, it can be stopped when all the questions of the type $q|q'$ have been asked. There will be at most $m \times (m - 1)$ such questions, and the actual number may be considerably less. In a recent application, involving 24 problems in high school algebra, the expert had to respond to 196 questions of the type $q|q'$, out of $24 \times 23 = 552$. This resulted in a knowledge space with 949 states, an impressive reduction because, at the start of the procedure, there were 2^{24} = 16,777,216 potential states. The procedure was pursued to the end, which further reduced the number of states to 312. The total time spent by the expert was approximately 7 hr.[2]

We are developing computer programs that will permit an efficient application of this procedure. The principle of these programs consists in maximizing the total number of pairs $q_1 \ldots q_n | q_{n+1}$ that could be added to either N_t or P_t as a result of asking a question, averaged over positive and negative responses. A sophisticated version of such a program looks ahead a certain number of steps. Because the program is interactive, the search for the best question to ask at Step t can end at the time of the response given at Step $t - 1$ (in a manner similar to chess-playing algorithms).

A Stochastic Learning Theory for a Knowledge Space

We shall refer to the procedure discussed in the last section as QUERY. Suppose that applying QUERY has resulted in the knowledge space specified by Equation 4 and represented in Figure 2. To assume that this knowledge space is necessarily correct in all of its details would be naive. Caution is advisable for several reasons. For one, an expert may not be fully consistent in the course of the questioning. Any realistic application of QUERY will take place over several days. Questions of Type Q2 asked the expert to be "practically certain." The interpretation of this phrase may vary from one day to the next. For another reason, if several experts participate, as is surely desirable, they will almost certainly give knowledge spaces that will differ in some respect, and these differences will have to be reconciled. The combination rule adopted, however reasonable it might be, may be a source of distortion.[3]

The knowledge space obtained from an application of QUERY has to be verified experimentally. If possible, a large sample of subjects from a specified population should be used. However, such data will unavoidably be noisy, and as such they cannot easily be confronted with a deterministic model such as that discussed in the preceding section. The appropriate step at this juncture is to construct a probabilistic model, which would specify the probabilities $\mathbb{P}(\mathbf{R})$ of all the 2^m possible patterns $\mathbf{R}$ of responses (with $m = |Q|$) by the equation

$$\mathbb{P}(\mathbf{R}) = \sum_{K \in \mathcal{H}} \mathbb{P}(\mathbf{R}|K)\mathbb{P}(K). \tag{8}$$

The assumptions of such a model must specify, for each pattern $\mathbf{R}$ and each state K (possibly up to the values of some parameters), the probability $\mathbb{P}(K)$ of state K in the population under consideration and the conditional probability $\mathbb{P}(\mathbf{R}|K)$ of observing $\mathbf{R}$ given K. This model can then be compared to the data by standard statistical techniques.

Falmagne (1989a) has developed a model of this type, and it is the topic of this section. An application of this model to some data pertaining to a high school mathematics test is described in the next section. We show how a test of the model against empirical data provides a way of evaluating the soundness of the knowledge space forming its core. A general strategy will be illustrated: Initially, we postulate a knowledge space containing many states, possibly too many, and designed to yield an acceptable fit. In a second stage, we make simplifying assumptions that lead to a gradual "pruning" of the knowledge space, with the succession of simpler and simpler submodels being tested by common likelihood ratio techniques.

Consider first the conditional probabilities $\mathbb{P}(\mathbf{R}|K)$ of the patterns of responses, given the states. Tightening our notation, we specify any pattern $\mathbf{R}$ of responses as a 0–1 vector with m components, where 0 and 1 represent the incorrect and correct responses, respectively, the order of the problems being fixed arbitrarily. Thus, in our standard example, ordering the problems by their numeric labels, we have

$$\mathbf{R} = (0, 1, 0, 1, 1), \tag{9}$$

denoting an error for Problems 1 and 3, and a correct response for the remaining three problems. To specify the conditional probabilities $\mathbb{P}(\mathbf{R}|K)$, we follow the traditional psychometric concept of "local independence" (Lazarsfeld, 1959). That is,

---

[2] We thank Maria Kambouri for serving as an expert and supplying us with these data.

[3] The problems of comparing and combining different knowledge spaces are bypassed here and will be discussed in a later publication.

conditional on the knowledge state of the subject, the responses to the problems are independent (i.e., any correlation between responses is completely explained by the states). Two parameters are associated with each problem: We write β_q for the probability of a "careless error" response to problem q. Thus, if problem q belongs to a subject's state, there is nevertheless a probability β_q that the subject will make an error in responding. On the other hand, at least in some situations (e.g., multiple-choice tests), there is a possiblility that a subject will provide a correct response to a problem not belonging to his or her state. We denote by ν_q the probability of a correct response to some problem q not belonging to the knowledge state. Writing $\mathbf{K}$ for the state of the subject, we have, using once more our standard example,

$$\mathbb{P}(\mathbf{R} = (0, 1, 0, 1, 1) | \mathbf{K} = \{3, 4\}) = (1 - \nu_1)\nu_2\beta_3(1 - \beta_4)\nu_5. \quad (10)$$

The left member denotes the probability that a subject in State $\{3, 4\}$ gives the pattern of responses of Equation 9. The product appearing in the right member illustrates the conditional independence of the responses inherent in the "local independence" principle.

We now turn to the second leg of the model, namely the specification of the probabilities $\mathbb{P}(K)$ of all the states. One brute-force possibility consists in simply assuming the existence of a probability distribution on the collection $\mathcal{K}$ of all the states. A major difficulty of this idea is that, even with a relatively small number of problems (say ≤ 30), the number of states will be very large, possibly on the order of several thousand. This automatically means that the model will have several thousand parameters, which is not an appealing prospect. Another drawback is that such a model offers no explanation for the specific differences, and also the similarities, among data from samples of subjects of different ages in the same population. In other words, even though two samples of different ages would, except for random variations, only differ by one number—the age of the group—the model would require reestimating all the parameters to account for the two sets of data. We shall return to this issue later in this article. The model that we are about to describe, which is much more ambitious, provides a stochastic description of the mastery of the successive items along the possible "learning paths." Within such a model, the difference between two age groups is represented simply by different values of one parameter, depending on the age of the group, and specifying the distribution of stages of learning.

We begin by strengthening the conditions defining a knowledge structure. We assume that any knowledge state is on a learning path, consisting in an increasing sequence of states, beginning with the empty state $\varnothing$ and finishing with the full set of questions Q, such that any state in the path different from $\varnothing$ contains exactly one more problem than the preceding one. Such a learning path is called a *gradation.* To avoid ambiguity, we reformulate this condition in more conventional mathematical terminology.

Definition 4: A chain of $|Q| + 1$ subsets of the set of problems Q in a knowledge structure $(Q, \mathcal{K})$, all of which are states, is called a gradation. If any state of the knowledge structure is contained in at least one gradation, the knowledge space is said to be *well graded.*

It is easy to see that any well-graded knowledge structure is discriminating (in the sense of Definition 1). As observed earlier, the knowledge space represented in Figure 2 is well graded and includes 16 gradations, which are represented below by the corresponding order in which the problems are mastered:

12345	12354	12435	13245
13254	13425	14235	14325
31245	31254	31425	32145
32154	32415	34125	34215.

To wit, for the last order in the list, Problem 3 is the first one to be mastered, followed by Problem 4, then Problems 2, 1, and finally 5. The corresponding gradation is the maximal chain of sets

$$\varnothing \subset \{3\} \subset \{3, 4\} \subset \{3, 4, 2\} \subset \{3, 4, 2, 1\} \subset \{3, 4, 2, 1, 5\}. \quad (11)$$

In general, we postulate the existence of a probability distribution on the set of all gradations, which will define, through additional features of the model, and via some parameters to be estimated from the data, a probability distribution on $\mathcal{K}$, the collection of states.

One could object that the number of gradations (a priori $|Q|!$) may be much larger than the number of states (a priori $2^{|Q|} < |Q|!$, for $|Q| > 3$); if so, such a model would be even less tractable than one postulating directly a probability distribution on $\mathcal{K}$. Our expectation was that, in practice, a small set of gradations would suffice to explain the data. This was confirmed by the application of the model described in the next section: Our analysis showed that a model with four gradations and nine states was capable of giving a good account (from a statistical viewpoint) of the test results of several hundred high school students.

The subject starts the process in State $\varnothing$, knowing nothing. Then, a gradation is chosen. We assume that gradation π is chosen with a probability equal to p_π. Thus,

$$\sum_{\pi \in G} p_\pi = 1,$$

in which G denotes the set of all gradations. (Thus, $|G| = 16$ in the knowledge space of Figure 2.) In general, the notation $\pi(1)$, $\pi(2)$, ..., represents the successive problems mastered along gradation π. If π denotes the gradation specified by Formula 11, we have $\pi(1) = 3$, $\pi(2) = 4, \ldots, \pi(5) = 5$. In the sequel, we consider a subject progressing along that gradation. It seems reasonable to suppose that the time required for mastering the first problem depends on the difficulty of that problem and on the learning ability, or learning rate, of the subject. We formalize this dual contribution as a random variable, which is denoted by $\mathbf{T}_{\pi(1),\lambda}$, where λ is a positive number representing the subject's learning rate and $\pi(1)$ denotes the first problem encountered in gradation π. Having mastered that problem, the subject is in state $\{\pi(1)\}$ and proceeds to master problem $\pi(2)$, the next one along that gradation; this will take a (random) time $\mathbf{T}_{\pi(2),\lambda}$. More generally, the time required to master the first j problems along gradation π is symbolized by the sum

$$\mathbf{T}_{\pi,j,\lambda} = \mathbf{T}_{\pi(1),\lambda} + \mathbf{T}_{\pi(2),\lambda} + \ldots + \mathbf{T}_{\pi(j),\lambda}.$$

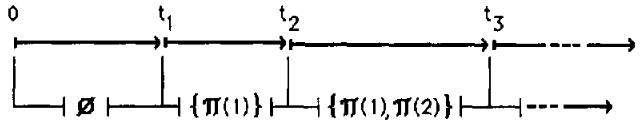

Figure 4. Succession of knowledge states at times 0, t_1, t_2, and so forth for a subject progressing along gradation π.

Notice that the learning rate is assumed to be constant for a given subject. A realization of this process is represented in Figure 4, in which the random variables take the values $\mathbf{T}_{\pi(1),\lambda} = t_1$, $\mathbf{T}_{\pi(2),\lambda} = t_2 - t_1, \ldots$. Between time 0 and time t_1, the subject remains in State $\varnothing$; between times t_1 and t_2, the subject's state is $\{\pi(1)\}$, and so forth. The probability that, at time τ, a subject with learning rate λ is in a state including the state $\{\pi(1), \ldots, \pi(j)\}$ is

$$\mathbb{P}(\mathbf{T}_{\pi(1),\lambda} + \ldots + \mathbf{T}_{\pi(j),\lambda} \leq \tau) = \mathbb{P}(\mathbf{T}_{\pi,j,\lambda} \leq \tau). \quad (12)$$

For $0 < j < m = |Q|$, the probability that at time τ this subject is exactly in state $\{\pi(1), \ldots, \pi(j)\}$ is

$$\mathbb{P}[\{\mathbf{T}_{\pi,j,\lambda} \leq \tau\} - \{\mathbf{T}_{\pi,j+1,\lambda} \leq \tau\}] = \mathbb{P}(\mathbf{T}_{\pi,j,\lambda} \leq \tau) - \mathbb{P}(\mathbf{T}_{\pi,j+1,\lambda} \leq \tau), \quad (13)$$

where the minus sign in the left member denotes the set difference between the event that the subject has mastered at least the first j problems in gradation π and the event that this subject has mastered at least $j + 1$ problems in that gradation. Equations 12 and 13 illustrate the type of computation inherent to this model. However, more work is required before reaching our goal, namely to obtain a specific prediction for the probability $\mathbb{P}(K)$ of state K. This probability clearly will vary with the time τ of the test, as is natural in a model tracing the course of learning. So far, we have considered a particular subject with learning rate λ. Actually, we assume that the learning rate has a distribution in the population under examination. We formalize this concept by a continuous random variable $\mathbf{L}$, with density f. The quantity λ in Equations 12 and 13 is a particular value of the random variable $\mathbf{L}$, and $f(\lambda)$ is the density of that value. We also introduce a random variable $\mathbf{G}$ for the gradation. That is, $\mathbf{G} = \pi$ denotes the event that a subject is channelled through gradation π. For simplicity, we suppose that the random variables $\mathbf{L}$ and $\mathbf{G}$ are independent. That is,

$$\mathbb{P}(\mathbf{L} \leq \lambda, \mathbf{G} = \pi) = \mathbb{P}\{\mathbf{L} \leq \lambda\}\mathbb{P}\{\mathbf{G} = \pi\} = \mathbb{P}(\mathbf{L} \leq \lambda\}p_\pi.$$

This assumption, however debatable it may be, cannot easily be avoided if one wishes the model to be tractable. Finally, $\mathbf{K}_\tau$ represents the state of the subject at time τ, and $\mathrm{G}(K)$ the set of all gradations passing through state K. Notice that for $j > 0$, we have necessarily

$$K = \{\pi(1), \ldots, \pi(j)\} \quad \text{if and only if} \quad |K| = j \text{ and } \pi \in \mathrm{G}(K).$$

Our apparatus is now complete, and the desired prediction can be worked out. We must compute the probability that a (randomly chosen) subject is in state K at time τ. This can be obtained as an average over all cases, that is, over all learning rates λ and all gradations π. Formally, we have

$$\mathbb{P}(\mathbf{K}_\tau = K) = \int_0^\infty \sum_{\pi\in\mathrm{G}(K)} \mathbb{P}(\mathbf{K}_\tau = K \mid \mathbf{L} = \lambda, \mathbf{G} = \pi)p_\pi f(\lambda)d\lambda. \quad (14)$$

We assume that the density f vanishes for $\lambda < 0$, which seems reasonable for such a concept as learning rate. Together with the local independence equation, illustrated by Equation 10, this is the fundamental equation of this model. Despite its rather formidable appearance, it actually leads to simple developments, under appropriate assumptions concerning the distributions of the random variables $\mathbf{L}$ and $\mathbf{T}_{i,\lambda}$. These are the only concepts remaining to be specified for the model to be applicable in a practical situation.

Indeed, when $|K| = j$, with $0 < j < m$, for instance, Equation 14 becomes

$$\begin{aligned}
&\mathbb{P}(\mathbf{K}_\tau = K) \\
&= \int_0^\infty \sum_{\pi\in\mathrm{G}(K)} \mathbb{P}(\mathbf{K}_\tau = \{\pi(1), \ldots, \pi(j)\} \mid \mathbf{L} = \lambda, \mathbf{G} = \pi)p_\pi f(\lambda)d\lambda \\
&= \int_0^\infty \sum_{\pi\in\mathrm{G}(K)} \mathbb{P}[(\sum_{i=1}^{j} \mathbf{T}_{\pi(i),\lambda} \leq \tau) - (\sum_{i=1}^{j+1} \mathbf{T}_{\pi(i),\lambda} \leq \tau)]p_\pi f(\lambda)d\lambda \\
&= \int_0^\infty \sum_{\pi\in\mathrm{G}(K)} [\mathbb{P}(\mathbf{T}_{\pi,j,\lambda} \leq \tau) - \mathbb{P}(\mathbf{T}_{\pi,j+1,\lambda} \leq \tau)]p_\pi f(\lambda)d\lambda. \quad (15)
\end{aligned}$$

Similar results hold for the two other cases, $K = \varnothing$ and $K = Q$.

We now turn to the parametric assumptions of this model. We assume that the random variables $\mathbf{T}_{i,\lambda}$ are independent and have the general gamma distribution (see Appendix), with parameters μ_i and λ, where μ_i is a measure of the difficulty of problem i. This means that the expectations and the variances of these random variables satisfy the relations

$$\mathrm{E}(\mathbf{T}_{i,\lambda}) = \mu_i/\lambda, \qquad \mathrm{Var}(\mathbf{T}_{i,\lambda}) = \mu_i/\lambda^2 \quad (16)$$

(Johnson & Kotz, 1970). The first equation in Display 16 states that the average time required to master problem i is the ratio of the difficulty μ_i of the problem by the learning rate λ of the subject. It follows that the total time $\mathbf{T}_{\pi,j,\lambda}$ required for a subject with learning rate λ to master the first j problems in gradation π is also distributed as a general gamma, with parameters

$$\mu_{\pi,j} = \sum_{i=1}^{j} \mu_{\pi(i)}, \quad (17)$$

and λ. From Equation 17, it is clear that

$$\mathrm{E}(\mathbf{T}_{\pi,j,\lambda}) = \mu_{\pi,j}/\lambda, \qquad \mathrm{Var}(\mathbf{T}_{\pi,j,\lambda}) = \mu_{\pi,j}/\lambda^2. \quad (18)$$

In words, the first equation in Display 18 means that the difficulties of the problems are additive: The average time required for a subject to master the first j problems along gradation π is the ratio of the sum of the difficulties of the problems, by the learning rate λ of the subject. It remains to specify the distribution of the learning rate in the population. It turns out that if we also assume that the learning rate random variable $\mathbf{L}$ has a general gamma distribution, with parameters $\psi > 0$ and $\delta > 0$, then Equation 15 can be evaluated by routine integration, leading to simple and explicit expressions in terms of the *incomplete beta function ratio*

$$\mathrm{I}_\theta(x, y) = \frac{1}{\mathrm{B}(x, y)} \int_0^\theta t^{x-1}(1-t)^{y-1}dt, \tag{19}$$

in which B is the *beta function* (see Appendix). We obtain, for $|K| = j, 0 < j < m$,

$$\mathbb{P}(\mathrm{K}_\tau = K) = \sum_{\pi \in G(K)} [\mathrm{I}_{\frac{\tau}{\tau+\delta}}(\mu_{\pi,j}, \psi) - \mathrm{I}_{\frac{\tau}{\tau+\delta}}(\mu_{\pi,j+1}, \psi)]p_\pi \tag{20}$$

and

$$\mathbb{P}(\mathrm{K}_\tau = Q) = \mathrm{I}_{\frac{\tau}{\tau+\delta}}(\mu_Q, \psi), \tag{21}$$

where μ_Q is defined by $\mu_Q = \mu_{\pi,m}$, for any gradation π. The third case is then obtained from the fact that

$$\mathbb{P}(\mathrm{K}_\tau = \varnothing) = 1 - \sum_{K \neq \varnothing} \mathbb{P}(\mathrm{K}_\tau = K).$$

The proof of these results are not given in this article (see Falmagne, 1989a).

The main appeal of these distribution assumptions lies in their mathematical convenience and in the fact that they lead to some reasonable interpretation of the parameters, as illustrated in Equations 16 and 18. However, these distribution assumptions are certainly not essential, and it is natural to ask whether they could lead to a rejection of an otherwise sound model. Notice, however, that neither the learning rate nor the times to mastery are observed directly. Our hope was thus that the model specified by the local independence assumption, and by Equations 20 and 21, would in fact be robust to the distribution assumptions—in the manner, for instance, that the analysis of variance model is robust to the assumption of a normal distribution of the errors. A successful test of this robustness is reported in the next section.

An Application of the Stochastic Learning Theory

The theory described in the preceding section was applied to the data provided by a sample of high school students from New York City. Our goals were twofold: We wanted, first, to test the theory on realistic data and second, to investigate whether it could be used, by likelihood ratio tests, to refine and simplify a hypothetical, well-graded knowledge space.

The subjects were 497 students, in Grades 10 or 11, all from the same high school in Manhattan. The test had 24 problems, which were completed by each student in the course of one class lasting 40 min. This amounts to less than 2 min per question, and some students may have been working under time pressure. Three equivalent versions of the test were used, which differed only in the order of the questions and in the specific numerical values of the quantities entering into the problems to be solved. Only the results for Problems 1–5 are considered herein. (A full analysis of a knowledge space for the 24 problems would require a much larger data set, that is, many more subjects.)

We started from a knowledge space slightly different from that represented in Figure 2. This knowledge space was derived from responses from an expert using methods similar to those described in the Building the Knowledge Space section. The data set was split randomly into two unequal sets, of 100 and 397 subjects.

In a first phase, the model was applied to the set of 100. This data set, even though much too small for statistical tests, nevertheless revealed some trends. This analysis was mostly based on heuristic considerations, the details of which are not reported herein. This led to the knowledge space of Figure 2, plus the hypothesis that the gradations 1 4 2 3 5, 1 4 3 2 5, and 3 4 2 1 5 had negligible probabilities.

All further analyses were based on the set of 397. Several statistical tests were performed, based on models of increasing simplicity. In each case, our goal was to predict the proportion of each of $2^5 = 32$ possible patterns of responses observed and to test the model by likelihood ratio test techniques.

The 13-Gradation Model

In principle, if we assume that the three gradations 1 4 2 3 5, 1 4 3 2 5, and 3 4 2 1 5 have zero probability and that the five "correct guessing" probabilities ν_i have also negligibly small values (which is reasonable in this case because the responses were open ended), the model is testable. Indeed, we have 31 degrees of freedom in the data and 24 parameters, which are listed below. Recall that τ is the time of the test, and ψ and δ are the parameters specifying the particular general gamma distribution of the learning rate.

12	probabilities p_π, corresponding to the 13 gradations
5	problem-difficulty parameters μ_i
5	careless-error probabilities β_i
1	parameter $\theta = \tau/(\tau + \delta)$
1	parameter ψ
24	total parameters.

Thus, in the rest of this section, we assume that $\nu_i = 0$ for $1 \leq i \leq 5$. Note that, as indicated by Equations 20 and 21, the parameters τ and δ are confounded in the equations of the model and are replaced by a single parameter $\theta = \tau/(\tau + \delta)$.

Actually, 397 subjects constitute too small a sample to yield reliable estimates of the 32 proportions (and approximate convergence of the likelihood ratio statistic to a chi-square random variable). The 32 patterns of responses were regrouped into 18, which only gives 17 degrees of freedom in the data. For the 13-gradation model to be testable, it has to be simplified in some way. We assumed (temporarily) that all the difficulty parameters μ_i have equal values, and that all the careless-error probabilities β_i are equal. Thus, 16 parameters remain, for 17 degrees of freedom. Because the simplifying assumptions are not very realistic, our goal was not so much to test the model as to estimate the gradation probabilities, with the hope that some of these estimates would prove to have negligibly small values, suggestive of a simpler model. We estimated the 16 parameters of the model by maximum likelihood, using the PRAXIS program (see Brent, 1973; Powell, 1964; we are indebted to Karl Gegenfurtner for making a C version of this program available to us). The key results are summarized in Table 4 and Figure 5. Table 4 gives the estimated probabilities of the 13 gradations, and Figure 5 displays the estimated probabilities of the states and of the transitions between states. From the gradation probabilities, we can derive the probabilities of the transitions between states. Indeed, the probability $P_{K,K+\{q\}}$ of mastering problem q imme-

Table 4
Estimated Probabilities (p_π) of the Gradations (π) in the 13-Gradation Model

π	p_π
1 3 2 4 5	.11
1 3 2 5 4	.06
1 3 4 2 5	.11
1 2 3 4 5	.00
1 2 3 5 4	.09
1 2 4 3 5	.00
1 4 2 3 5	Set to 0
1 4 3 2 5	Set to 0
3 2 4 1 5	.00
3 2 1 4 5	.11
3 2 1 5 4	.01
3 4 1 2 5	.03
3 4 2 1 5	Set to 0
3 1 2 4 5	.45
3 1 2 5 4	.03
3 1 4 2 5	.00

diately after mastering all the problems in state K is the conditional probability of a gradation going through $K + \{q\}$, given that it passes through K. By the rules for conditional probabilities, this equals the ratio of the probability for a gradation passing through both K and $K + \{q\}$ to that for a gradation passing through K only:

$$P_{K,K+\{q\}} = [\sum_{\pi\in G(K)\cap G(K+\{q\})} p_\pi]/[\sum_{\pi\in G(K)} p_\pi].$$

For instance, the estimated probability $P_{\varnothing,\{3\}}$ of a transition from the null state to State $\{3\}$ is equal to the sum of the estimated probabilities of all the gradations π such that $\pi(1) = 3$, namely, the sum of all the probabilities in the last column of the table: $.11 + .01 + .03 + .45 + .03 = .63$. Similarly, the estimated probability $P_{\{3\},\{3,2\}}$ of a transition from State $\{3\}$ to State $\{2, 3\}$ is equal to $(.11 + .01)/.63 = .19$. All the transition probabilities are given in Figure 5. The state probabilities are computed from Equations 20 and 21. The log likelihood ratio statistic (testing the 13-gradation model against the multinomial model with 18 classes) is, asymptotically, chi-square distributed with 1 degree of freedom. Even though the model is strongly rejected (chi-square statistic = 11.3), the results suggest that some states may never be vistited and could possibly be removed. In particular, States $\{3, 4\}$, $\{1, 2, 4\}$, $\{2, 3, 4\}$, and perhaps $\{1, 3, 4\}$ and $\{1, 2, 3, 5\}$ may have a negligible probability. This can be seen in Figure 5.

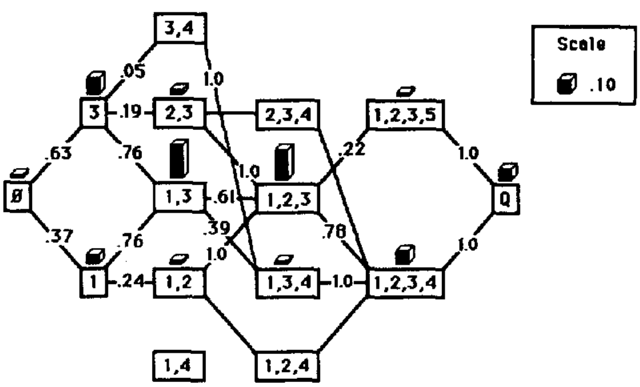

Figure 5. The 13-gradation model with the estimated probabilities of the knowledge states (boxes) and of the transitions (lines) between states. (The gradations 14235, 14325, and 34215 have been assigned probability zero a priori. The corresponding segments are omitted in the graph.)

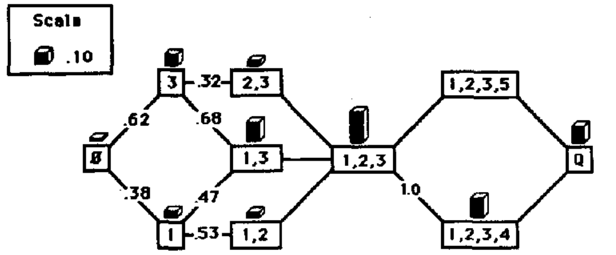

Figure 6. The eight-gradation model with the estimated probabilities of the knowledge states (boxes) and of the transitions (lines) between them. (The states $\{1, 4\}$, $\{3, 4\}$, $\{1, 2, 4\}$, $\{1, 3, 4\}$, and $\{2, 3, 4\}$ have been dropped and are not displayed.)

The 8-Gradation Model

In the next knowledge space that we tested, the eight gradations 1 2 4 3 5, 1 3 4 2 5, 1 4 2 3 5, 1 4 3 2 5, 3 2 4 1 5, 3 4 1 2 5, 3 4 2 1 5, and 3 1 4 2 5 were assigned zero probability a priori. In applying the model, all parameters are free to vary, except that the values of the difficulty parameter μ_i are assumed to be equal. Thus, we have $17 - (7 + 1 + 5 + 2) = 2$ degrees of freedom. As stated earlier, the parameters are estimated by maximizing the log likelihood using PRAXIS. The fit of the model is considerably better: The log likelihood ratio statistic is 3.9 for 2 degrees of freedom. (Again, the model is tested against the multinomial model with 18 classes.) The results are summarized in Figure 6. The estimated probability of State $\{1, 2, 3, 5\}$ appears to be very low. This state is dropped in our next model.

The 4-Gradation Model

In this model, only four gradations remain: 3 2 1 4 5, 3 1 2 4 5, 1 3 2 4 5, and 1 2 3 4 5. All parameters are free to vary, which gives $17 - (3 + 5 + 5 + 2) = 2$ degrees of freedom. A good fit is obtained: Chi-square statistic = 3.5 for 2 degrees of freedom. (As before, the model is tested against the multinomial model.) The results are summarized in Figure 7 and Table 5. The numerical values obtained for the parameters and for the state probabilities prompt several remarks.

The values of the careless-error probabilities β_i seem high. We recall, however, that the subjects had only 40 min for a 24-problem test, and some subjects may have had to work under time pressure in the last part of the test (which was not the same for all subjects, because three equivalent versions of the test

Table 5
Estimated Values of the Parameters in the 4-Gradation Model

Parameter	Estimate	Parameter	Estimate	Parameter	Estimate
1 3 2 4 5	.19	β_1	.15	μ_1	2.03
1 2 3 4 5	.22	β_2	.37	μ_2	4.85
3 2 1 4 5	.33	β_3	.13	μ_3	3.16
3 1 2 4 5	.26	β_4	.31	μ_4	4.59
θ	.71	β_5	.44	μ_5	2.58
ψ	4.72				

Note. The first four entries in the second column are the estimated probabilities of the four gradations in the first column. The chi-square statistic is 3.5, for 2 degrees of freedom.

were used). This supposition is supported by an examination of the response sheets, which showed that the last pages of the test were sometimes incomplete. (This could be checked because the student had to mark each problem attempted.)

The quantity θ is a compound parameter, $\theta = \tau/(\tau + \delta)$, where τ is the time of the test (which may be measured by the grade level) and δ is a parameter of the learning rate distribution, inversely proportional to the expectation. This means that a large value of θ (that is, a value close to 1) may indicate a large value of the mean learning rate or the fact that the students are tested late in their education curriculum. A somewhat better understanding of this issue could be obtained in a situation in which several age groups from the same population are compared. This situation is illustrated by the computations at the end of this section. It may perhaps come as a surprise that, as shown in Figure 7, the estimated probability of State $\{1, 2, 3, 4\}$ is smaller than that of the two adjacent states, $\{1, 2, 3\}$ and $Q = \{1, 2, 3, 4, 5\}$. This effect is probably due to the fact that state Q is the last possible state in the learning process and may act as a buffer. In other words, there may be some students in our sample whose competence is beyond state Q. In our representation, all such students are gathered in state Q for lack of a more advanced state in the knowledge space. (A similar remark applies to Figure 8.)

Simpler Models

The analysis of the four-gradation model is not suggestive of further simplifications. We nevertheless tested several simpler models. We began by systematically testing all three-gradation submodels of the four-gradation model. In all cases, the test statistic was the log likelihood ratio of the three-gradation model against the four-gradation model of Figure 7. Thus, the test statistic is, asymptotically, chi-square distributed with 1 degree of freedom. Only one of the four submodels gave a significant result (at the 5% level), namely the model assigning a priori zero probability to the gradation 1 2 3 4 5 (chi-square statistic = 6.82 for 1 degree of freedom). It is probably safe to conclude that our data were not rich enough to distinguish between the other submodels.

A similar situation arose with the two-gradation models. We tested all two-gradation submodels of the four-gradation model against the best fitting three-gradation model in which they are included. In three of six cases, nonsignificant chi-squares were obtained. These models are listed in Table 6 with their chi-square statistics.

Finally, we tested all 4 one-gradation models, each time against the best fitting two-gradation model in which they were included (for which a nonsignificant chi-square had been obtained). All were strongly rejected. The results are summarized in Table 7. Notice that the one-gradation model 1 3 2 4 5 gives a better fit than the two-gradation model (1 3 2 4 5, 1 2 3 4 5), which includes it. The paradox is explained by the fact that the denominators in the likelihood ratio statistic are different. That is, these models are tested against different alternatives.

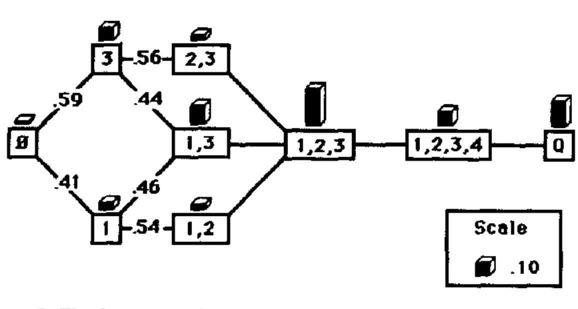

Figure 7. The four-gradation model, with the estimated probabilities of the knowledge states (boxes) and of the transitions (lines) between them.

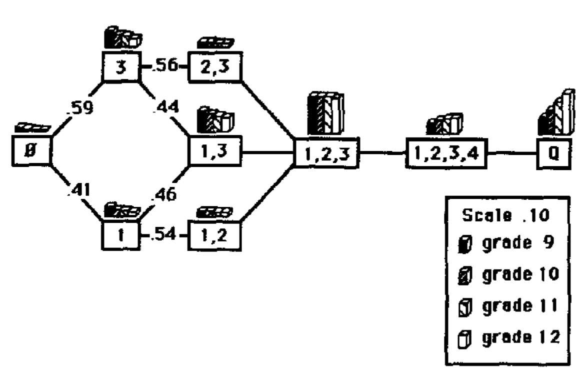

Figure 8. The four-gradation model. (Probabilities of occupation of the knowledge states have been computed for four age groups. The parameter values are those of Table 5, except for θ.)

Table 6
The Six 2-Gradation Models, With Their Chi-Square Statistics

Model	Chi-square statistic (1 *df*)
1 3 2 4 5, 1 2 3 4 5	19.5
3 1 2 4 5, 3 2 1 4 5	12.1
1 2 3 4 5, 3 2 1 4 5	7.5
1 3 2 4 5, 3 2 1 4 5	3.3
1 3 2 4 5, 3 1 2 4 5	3.1
1 2 3 4 5, 3 1 2 4 5	2.5

Note. A chi-square statistic of 3.843 is significant at the 5% level.

Other Distributional Assumptions

The main purpose of the learning model was to provide a practical tool for testing and refining some hypothetical knowledge space. From this standpoint, the role of the distributional assumptions has to be examined with care. The general gamma distributions for the learning rate and for the learning latencies were an appealing choice for several reasons. For instance, the learning latencies were additive over successive problems, with the difficulty parameter for a set of successive problems being the sum of its components (cf. Equation 17). These assumptions also led to simple mathematical expressions for our predictions, which render the model easily applicable. (The integrals in equations such as Equation 15, compounding the general gamma for the learning latencies with the general gamma for the learning rate, could be evaluated explicitly, leading to the incomplete beta function ratio expressions in Equations 20 and 21.) Nevertheless, these are considerations of convenience, and it would be a serious drawback if we were to discover that our conclusions concerning the knowledge space depend strongly on these distributional assumptions.

We did not investigate this question fully. We did, however, test the four-gradation model represented in Figure 7 with different distributional assumptions. The results were encouraging.

Rather than adopting other distributions for the learning rate and for the learning latencies, we took the easier tack of modifying directly the distributions entering into Equations 20 and 21. Specifically, we replaced in these equations the incomplete beta function ratio

$$\theta \rightarrow \mathrm{I}_\theta(\mu_K, \psi)$$

with the distribution function

$$\theta \rightarrow \Phi\left[\frac{\log\frac{\theta}{1-\theta} - \mu_K}{\sigma}\right],$$

where Φ is the distribution function of the standard normal random variable; $0 < \theta < 1$; and $\mu_K = \sum_{i \in K} \mu_i$ and σ are parameters, with $\sigma > 0$. Notice that both of these expressions define distribution functions on the interval (0, 1).

The fit of the model was as good as that obtained for the general gamma (the chi-square value was actually the same: 3.5 for 2 degrees of freedom), and the values obtained for the estimated parameters common to both models were remarkably close. We may thus tentatively conclude that the learning model is, as we had hoped, fairly robust to the distributional assumptions.

Predicting Data for Several Age Groups

One particular feature of the learning model deserves to be emphasized. Consider some hypothetical data obtained from several samples of subjects (from the same population) of different age groups, say from Grades 9–12. Even the most pessimistic observer of the educational process would expect the test results for these samples to differ widely. Predicting the overall data by a single model is somewhat of a challenge. In fact, the learning model discussed here provides simple and specific predictions for such a situation. Indeed, the only difference between the samples is the time τ of the test, which enters into the expression of the incomplete beta function ratio in Equations 20 and 21. All the other parameters of the model must remain the same if the samples come from the same population. Changing the time τ of the test, for example setting τ = 9, 10, 11, or 12, should, in principle, suffice to yield the desired predictions.

To illustrate the possibilities of such an analysis, we made some comparative computations involving the four-gradation model of Figure 7. As indicated above, we considered Grades 9–12. All the parameter values used in the computations were those listed in Table 5, except for θ, which was different in each of the four age groups. We set $\delta = 4.13$ to obtain

$$\theta_{10} = \frac{10}{10+\delta} = \frac{10}{10+4.13} = .71,$$

as in Table 5. The values of θ for the other age groups were then obtained from

$$\theta_9 = \frac{9}{9+4.13} = .69,$$

$$\theta_{11} = \frac{11}{11+4.13} = .73,$$

$$\theta_{12} = \frac{12}{12+4.13} = .74.$$

(All numbers are rounded to the second decimal.) Using these values of θ and the other values of the parameters in Table 5, we computed the probabilities of occupations of the various knowledge states. The results are presented in Figure 8, which is similar to Figure 7, except that four different age groups are

Table 7
The Four 1-Gradation Models, With Their Chi-Square Statistics

Model	Chi-square statistic (1 *df*)
1 3 2 4 5	16.9
1 2 3 4 5	32.3
3 2 1 4 5	20.5
3 1 2 4 5	12.0

Table 8
Illustration of a Deterministic Procedure

		Knowledge state								
Problem	Response	∅	{1}	{3}	{1, 3}	{2, 3}	{1, 2}	{1, 2, 3}	{1, 2, 3, 4}	{1, 2, 3, 4, 5}
2	F	*	*	*	*	—	—	—	—	—
1	C	—	*	—	*	—	—	—	—	—
3	C	—	—	—	*	—	—	—	—	—

Note. Asterisks indicate the possible states remaining after the given response; dashes indicate the states discarded. F = incorrect response; C = correct response.

represented. Such results suggest that the learning model may be useful as a tool for monitoring educational progress.

Uncovering the State of a Student: Two Markovian Procedures

Suppose that, having applied the techniques described in the preceding sections, we have constructed and verified a particular well-graded knowledge space. We now turn to the final goal of our project, namely to design efficient knowledge assessment procedures. Before computer programs are written for that purpose, it is advisable to analyze some possible fundamental procedures.

Consider the knowledge space obtained for the five problems as a result of the analysis presented in the previous section (cf. Figure 7):

$$\mathscr{K} = \{\varnothing, \{1\}, \{3\}, \{1, 3\}, \{2, 3\}, \{1, 2\}, \{1, 2, 3\}, \{1, 2, 3, 4\}, \{1, 2, 3, 4, 5\}\}. \quad (22)$$

This knowledge space is well graded (with four gradations). We begin by assuming the subject's responses never result from a careless error or a lucky guess (this assumption is relaxed later). Consider an assessment procedure in which the first problem posed is Problem 2. If an incorrect response is obtained, all states containing 2 must be discarded. We indicate this conclusion by marking (with an asterisk) the remaining possible states, as indicated, in the second line of Table 8. (Obviously, if a correct response had been observed, all the other states on that line would have been marked.) Next, if Problem 1 is presented and we observe a correct response, only $\{1\}$ and $\{1, 3\}$ remain as possible states. Finally, asking Problem 3 and observing a correct response results in State $\{1, 3\}$. This is the unique state consistent (in this deterministic framework) with the data: 2 failed, 1 and 3 correct. Clearly, every state can be uncovered by such a procedure, which can be represented by a binary decision tree (see Figure 9).

Several issues pertaining to such a formalization were explored by Degreef, Doignon, Ducamp, and Falmagne (1986). For example, suppose that a teacher describes just one particular decision tree, of the kind exemplified in Figure 9, to be used in testing students regarding a specific body of knowledge. Is this information sufficient to recover the full collection of knowledge states? The answer is yes, provided that the knowledge structure is partially ordinal in the sense of Definition 1. In the general case (when no constraints are placed on the knowledge structure), if all the decision trees are known, then the collection of all states is recoverable.

A serious difficulty for this assessment procedure is that it cannot deal effectively with the inherent randomness, or more generally, instability, of a subject's performance. Obvious examples are the careless errors and lucky guesses formalized in the model discussed in the two preceding sections. The instability may also have a more fundamental origin: The subject's knowledge state may actually change slightly in the course of the procedure. This might happen, for instance, in the case of a subject tested on material that had been learned a long time before the test and not exercised recently. The early part of the test may facilitate the retrieval of material relevant to the last part. In any event, more robust procedures are required, which would be capable, despite the noisy data, of uncovering exactly, or at least approaching closely, the knowledge state of a student.[4]

Two classes of Markovian procedures have been investigated by Falmagne and Doignon (1988a, 1988b). Both of these classes can be represented in a common framework, as illustrated by Figure 10. At the onset of trial n of the procedure, the information gathered from the responses to problems presented previously is summarized by a *plausibility function*, which assigns plausibility values to all the states. The value of this function is used to choose the problem asked on that trial, using a *questioning rule*, which may take various forms. The subject's response is then observed, and it is assumed to be governed by the subject's knowledge state through a *response rule*. In the simplest case, we assume that the response is correct if and only if the problem belongs to the subject's knowledge state. We may also introduce parameters for careless errors or lucky guesses, as exemplified by the local independence (Equation 10). Finally, the plausibility function is recomputed by an *updating rule*, initiating trial $n + 1$.

From a mathematical standpoint, the problem consists in investigating conditions under which the subject's knowledge state can be uncovered through systematic testing following a given procedure. A more complex situation in which the subject's state varies over trials is also considered. In particular, we assume that there exists a probability distribution on the

[4] We are assuming, of course, that the knowledge state of the student changes at most once or twice in the course of the procedure.

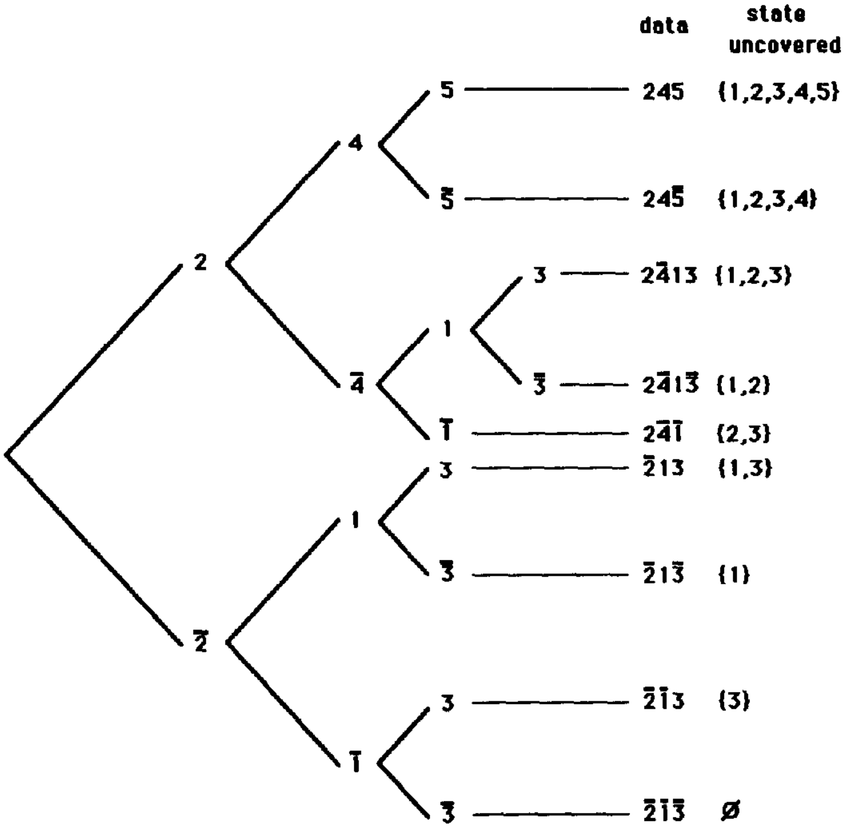

Figure 9. Example of a binary decision tree that permits the discovery of any state in the knowledge space described by Equation 22 and Figure 7. (The overbars indicate incorrect responses.)

collection of all the knowledge states. This probability distribution, referred to as a *stochastic (knowledge) state,* is not observable directly, but it governs the subject's manifest behavior. The goal is then to uncover this probability distribution, or at least some of its important characteristics.

In the next two subsections, the two classes of procedures and some of our main results are outlined. (For a complete description, see Falmagne & Doignon, 1988a, 1988b.) Except when mentioned otherwise, we assume that the subject is in some state K_0, constant over trials.

A Discrete Markov Chain Procedure

For concreteness, only a special case of the theory developed by Falmagne and Doignon (1988b) is described, one that appears important for practical use. As in Table 8, the plausibility function implements a marking of the states considered plausible at a given trial. This procedure is designed to quickly narrow down the collection of marked states until only one such state remains. At that stage, because of the noisy components of the situation, the remaining state may not coincide exactly with the subject's state but may reasonably be assumed to have much in common with it. Further questioning under a slightly altered rule achieves a final refinement of the assessment. More specifically, we denote by $\mathbf{M}_n$ the set of marked states on trial n. The problem asked on trial n, denoted by $\mathbf{Q}_n$, is chosen to discriminate between the marked states. At the outset, all states are marked. In our example, this means that $\mathbf{M}_1 = \mathscr{H}$, with $\mathscr{H}$ as in Equation 22. The successive problems $\mathbf{Q}_1$, $\mathbf{Q}_2$, and so forth, should be chosen to discriminate quickly among the states in $\mathbf{M}_1$. Suppose that $\mathbf{Q}_1 = 5$. The knowledge space $\mathscr{H} = \mathbf{M}_1$ has nine states, only one of which contains Problem 5. If the response is correct and assuming that the probability of a correct guess is negligible (we shall return to the issue of the "noise" in a moment), the subject's state is known. However, if the response is incorrect, eight states remain possible, and we have learned little. A priori, all states are equiprobable, so the probability of learning little is relatively high. The situation seems intuitively better if we set $\mathbf{Q}_1 = 1$ or $\mathbf{Q}_1 = 3$: In either case, there are six states of $\mathbf{M}_1$ containing $\mathbf{Q}_1$ and three states not containing $\mathbf{Q}_1$. Whether or not the subject's response is correct, a substantial reduction of uncertainty will take place. The best choice for $\mathbf{Q}_1$ seems to be Problem 2, which splits $\mathbf{M}_1$ into four states containing it and five states not containing it. Table 9 gives the analysis for all five problems. In general, whenever $\mathbf{M}_n$ contains more than one state, the questioning rule is formulated as follows:

For $|\mathbf{M}_n| > 1$, choose $\mathbf{Q}_n$ such that $\mathbf{M}_n$ is split as equally as possible between the subset of all states in $\mathbf{M}_n$ containing $\mathbf{Q}_n$, and the subset

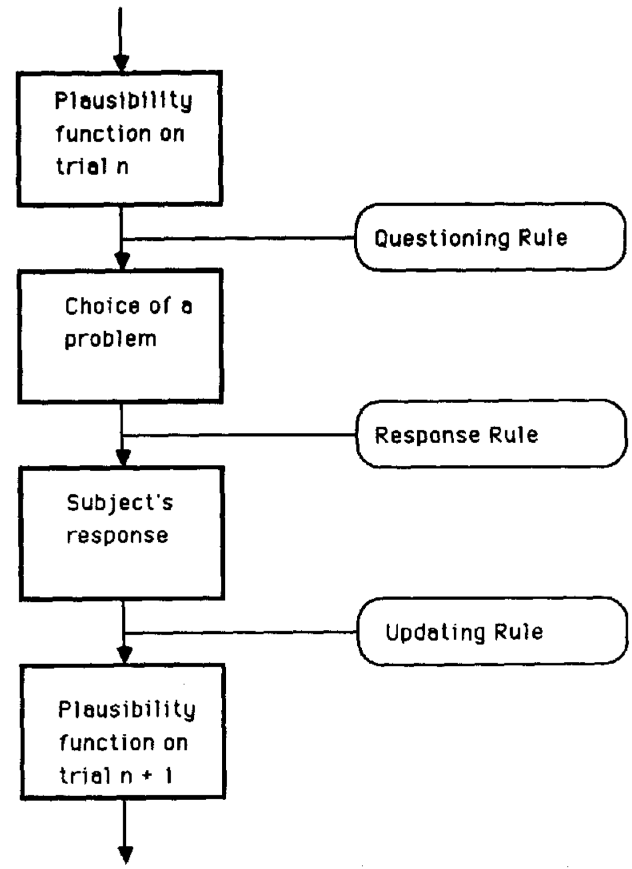

Figure 10. Diagram of the two classes of Markovian assessment procedures.

> of all states in $\mathbf{M}_n$ not containing $\mathbf{Q}_n$. If there are several problems satisfying these conditions, choose randomly among them, with equal probabilities.

We recall that $|S|$ denotes the number of elements in a set S. According to this rule, Problem 2 should be the first problem asked. If an error is observed, the second problem should be chosen randomly between 1 and 3. The updating rule is then

> If the response is correct, set $\mathbf{M}_{n+1}$ equal to the subset of $\mathbf{M}_n$ containing all the states K such that $\mathbf{Q}_n \in K$. Otherwise, set $\mathbf{M}_{n+1}$ equal to the subset of $\mathbf{M}_n$ containing all those states K such that $\mathbf{Q}_n \notin K$.

We denote by $\mathbf{R}_n = 1$ and $\mathbf{R}_n = 0$ a correct and incorrect response on trial n, respectively. The updating rule is defined by the pair of equations

$$\mathbf{M}_{n+1} = \begin{cases} \{K \in \mathbf{M}_n \mid \mathbf{Q}_n \in K\} & \text{if } \mathbf{R}_n = 1; \\ \{K \in \mathbf{M}_n \mid \mathbf{Q}_n \notin K\} & \text{if } \mathbf{R}_n = 0. \end{cases}$$

Clearly, successively applying these two rules will result in a set of marked states containing exactly one state, say $\mathbf{M}_k = \{K\}$. As was noted, this state need not be taken as the final assessment, and further questioning usually is required. The next problem asked is based on $\mathbf{M}_k$ augmented with its *neighboring states,* which are defined as those states differing from K by exactly one problem; in other terms, states of the form $K + \{q\}$ or $K - \{q\}$. Table 8 can actually be taken as representing the first three trials of an application of the two rules stated above. On Trial 3 + 1, the single state $\{1, 3\}$ remains. Its neighboring states are $\{1\}$, $\{3\}$, and $\{1, 2, 3\}$. We denote by $\mathbf{M}'_n$ the collection of states containing the unique set in $\mathbf{M}_n$ together with its neighboring states. Thus, if $\mathbf{M}_4 = \{1, 3\}$, then

$$\mathbf{M}'_4 = \{\{1, 3\}, \{1\}, \{3\}, \{1, 2, 3\}\}. \tag{23}$$

The modified questioning rule states

> For $|\mathbf{M}_n| = 1$, choose $\mathbf{Q}_n$ such that $\mathbf{M}'_n$ is split as equally as possible between the subset of all states in $\mathbf{M}'_n$ containing $\mathbf{Q}_n$, and the subset of all states in $\mathbf{M}'_n$ not containing $\mathbf{Q}_n$. If there are several problems satisfying these conditions, choose randomly among them, with equal probabilities.

The only difference between this questioning rule and the preceding one is that the split applies to $\mathbf{M}'_n$ rather than to $\mathbf{M}_n$. Applying this questioning rule to $\mathbf{M}'_4$ in Equation 23, we see that $\mathbf{Q}_4$ should be chosen with equal probabilities between Problem 1, 2, or 3, because each of them splits $\mathbf{M}'_4$ into two subsets, one of which contains one state and the other, three states. Suppose that Problem 2 is chosen (thus, $\mathbf{Q}_4 = 2$) and the response is correct. We have, then, $\mathbf{M}_5 = \{\{1, 2, 3\}\}$. This illustrates the updating rule in the case $|\mathbf{M}_n| = 1$, which is as follows:

> Suppose that $\mathbf{M}_n = \{K\}$ and $\mathbf{Q}_n = q$:
>
> i. If $q \in K$ and the response is correct, set $\mathbf{M}_{n+1} = \{K\}$
> ii. If $q \in K$ and the response is incorrect, set $\mathbf{M}_{n+1} = \{K - \{q\}\}$
> iii. If $q \notin K$ and the response is correct, set $\mathbf{M}_{n+1} = \{K + \{q\}\}$
> iv. If $q \notin K$ and the response is incorrect, set $\mathbf{M}_{n+1} = \{K\}$.

Thus, $|\mathbf{M}_n| = 1$ necessarily implies $|\mathbf{M}_{n+1}| = 1$.

The mathematical properties of this procedure have been analyzed in detail by Falmagne and Doignon (1988b). The summary of our results given below is intended for a reader familiar with the basic concepts of Markov chains theory (Feller, 1970; Kemeny & Snell, 1965). If the subject is in some knowledge state K_0, constant over trials, then the stochastic process $(\mathbf{M}_n)$ is a Markov chain. This is also true for the processes $(\mathbf{M}_n, \mathbf{Q}_n)$ and $(\mathbf{M}_n, \mathbf{Q}_n, \mathbf{R}_n)$. We showed that if there are no careless errors and no correct guessing, then $\mathbf{M}_n$ converges to the subject's state K_0 in the sense that

$$\lim_{n \to \infty} \mathbb{P}\{\mathbf{M}_n = \{K_0\}\} = 1.$$

We also proved that if there is a positive probability of some careless errors but still no correct guessing, then the Markov chain $(\mathbf{M}_n)$ had a unique ergodic set E containing $\{K_0\}$, and possibly some Markov states $\{K\}$ with $K \subseteq K_0$, but no other Markov states. If, in addition, $\beta_q > 0$ for all $q \in K_0$, then E is in fact the family of all those Markov states $\{K\}$ such that $K \subseteq K_0$.

In the case where the subject's state is stochastic, the process $(\mathbf{M}_n)$ remains Markovian. It was found that, in certain situations, the distribution characterizing the subject's stochastic state could be estimated from the frequencies of the observable

Table 9
Illustration of Questioning Rule in the Discrete Markov Chain Procedure

	Knowledge state								
Problem	$\varnothing$	{1}	{3}	{1, 3}	{2, 3}	{1, 2}	{1, 2, 3}	{1, 2, 3, 4}	{1, 2, 3, 4, 5}
2	−	−	−	−	+	+	+	+	+
1	−	+	−	+	−	+	+	+	+
3	−	−	+	+	+	−	+	+	+
4	−	−	−	−	−	−	−	+	+
5	−	−	−	−	−	−	−	−	+

Note. Each problem splits the set of states into two subsets containing, respectively, the states containing the problem (+) and the states not containing the problem (−). The best discrimination is achieved by Problem 2, which splits $\mathbf{M}_1$ into two subsets of five and four states, respectively.

responses. For the time being, however, these results have only anecdotical interest.

Some preliminary simulations were performed (see Villano et al., 1987) that suggest that this discrete Markov procedure may be as efficient as the considerably more demanding one outlined in the next subsection.

A Continuous Markov Procedure

Only a short description of the stochastic process developed by Falmagne and Doignon (1988a) will be given. We assume that the plausibility function takes the form of a likelihood function $K \to L_{n,K}$ defined, for any trial number n, on the collection $\mathcal{K}$ of all knowledge states. We suppose that

$$0 < L_{n,K} < 1 \quad \text{for all states } K \in \mathcal{K}, \quad \text{and} \quad \sum_{K \in \mathcal{K}} L_{n,K} = 1.$$

Thus, L_n is a probability distribution on $\mathcal{K}$, assigning, on every trial, a positive probability to every knowledge state.

Two questioning rules were considered. The first one, called the *half-split rule,* is similar in spirit to the questioning rule used in the discrete Markov procedure just described. Notice that each problem q splits the total mass of the likelihood function into two parts:

$$\rho_n(q) = \sum_{K \in \mathcal{K}_q} L_{n,K} \quad \text{and} \quad 1 - \rho_n(q) = \sum_{K \notin \mathcal{K}_q} L_{n,K},$$

that is, the total mass of all those states containing q and the total mass of all those states not containing q. Problems for which these two masses are as close as possible may be considered more informative than others. The problem $\mathbf{Q}_n$ is selected at random among such questions.

A more elaborate way of evaluating the information content of a problem relies on a computation of the entropy of the likelihood function. On trial n, the entropy is by definition

$$\mathrm{H}(L_n) = -\sum_{K \in \mathcal{K}} L_{n,K} \log_2 L_{n,K}.$$

The idea is to choose $\mathbf{Q}_n$ so as to minimize the entropy of the likelihood function on trial $n + 1$. By itself, however, this concept does not provide a manageable criterion for choosing $\mathbf{Q}_n$ because this is actually an expected entropy, the value of which depends, for any problem q, on the probability $\rho_n(q)$ of a correct response and also on the particular updating rule adopted. In other terms, to apply this rule, we must compute, for each problem q, the conditional expected entropy

$$\mathrm{H}(L_{n+1} \mid L_n, \mathbf{Q}_n = q) = \rho_n(q)\mathrm{H}(L_{n+1} \mid L_n, \mathbf{Q}_n = q, \mathbf{R}_n = 1) + [1 - \rho_n(q)]\mathrm{H}(L_{n+1} \mid L_n, \mathbf{Q}_n = q, \mathbf{R}_n = 0) \quad (24)$$

and minimize $\mathrm{H}(L_{n+1} \mid L_n, Q_n = q)$ over all problems q. Any questioning rule based on this principle is called *informative.* Equation 24 may also be written more explicitly

$$\mathrm{H}(L_{n+1} \mid \mathbf{Q}_n = q) = \rho_n(q)\mathrm{H}[u(L_n; q, 1)] + [1 - \rho_n(q)]\mathrm{H}[u(L_n; q, 0)]$$

in terms of the updating rule u:

$$(L_n, \mathbf{Q}_n, \mathbf{R}_n) \to u(L_n; \mathbf{Q}_n, \mathbf{R}_n) = L_{n+1}.$$

The updating rule u is thus an operator depending on the problem asked and the response given (and on some numerical parameters to be calibrated), which transforms the likelihood function on trial n into the likelihood function on trial $n + 1$. Two classes of updating rules are analyzed in Falmagne and Doignon (1988a). One of them is shown to be "commutative," in the sense that it results in the same value for the likelihood function on trial n, regardless of the order of the problems asked on preceding trials. This commutative rule is proven to be consistent with a Bayesian principle. (For a survey of similar updating rules, in this and other contexts, see Landy & Hummel, 1986.)

As mentioned earlier, despite their apparently much more sophisticated use of the response data, these continuous Markov procedures do not seem to be preferable, from the viewpoint of the accuracy of the assessment, to the discrete Markov procedure outlined in the first part of this section. However, our results in this regard are far from complete.

Discussion

With this article, we intended to give a comprehensive description of our progress in a fairly ambitious project, namely, designing and implementing efficient knowledge assessment procedures that are capable of competing successfully with a competent human examiner engaged in a one-to-one interaction with a student or a candidate and general enough to be widely applicable. Such a project necessarily involves many phases. The puzzle is not complete, but there are relatively few missing pieces, and it is quite clear what they are. There are also several potential objections to our approach, which we consider in this section.

The algebraic core of the theory consists in the concept of a knowledge space, conceived as a collection of subsets of problems or questions. Each of those subsets is a knowledge state, which we interpret as the set of all problems that a subject is capable of solving correctly, without help or guessing, with the assumption that careless errors do not occur.

A potential criticism to this formalization concerns the coding of a subject's response into two categories only: correct or incorrect. In many situations, other types of responses are recorded, such as a response latency or the method of solution used. In our view, this binary coding is not a serious limitation. Suppose, for instance, that the latency of the reaction to a particular problem has been measured, and that we want to include this information in the description of the subject's response. Thus, the set of possible responses to this problem is theoretically uncountable. In practice, however, the data can always be recoded to fit our framework. Because the assessor is certainly not interested in recording the response latency in milliseconds, or even in seconds, we can always decompose the particular problem into a finite list of related problems with binary responses, such as a correct response in less than 30 s (yes or no), a correct response in less than 1 min (yes or no), and so forth. The theory could be reworked to allow for multiple responses if this generalization is judged important.

We also reconsider the $\cup$-closure axiom, that is, the assumption that the union of knowledge states is always a knowledge state. This axiom plays an important role in our construction. Nevertheless, examples can be given that suggest that this axiom may not always be valid in practical applications. One example involves two world-class mathematicians working in different fields. Suppose that their respective knowledge states are K and K'. It is most unlikely that there would be another mathematician whose knowledge state would be equal to or would include $K \cup K'$. Thus, the $\cup$-closure axiom may be adding useless states at the "edge" of the knowledge structure. Another counterexample (provided by a reviewer) involves the concept of skill described in the Knowledge Spaces and Skills section. Consider the case of a set $Q = \{1, 2, 3\}$, with four relevant skills—**a**, **b**, **c**, and **d**—and the following correspondence:

Problem	Skills required
1	**a, b**
2	**c, d**
3	**a, c**

If we assume that an individual can have any subset of skills, the feasible knowledge states are: $\varnothing$, $\{1\}$, $\{2\}$, $\{3\}$, $\{1, 3\}$, $\{2, 3\}$, and $\{1, 2, 3\}$. Indeed, a single skill is never sufficient to solve a problem. Similarly, some sets of two skills are also useless, namely $\{\mathbf{a}, \mathbf{d}\}$, $\{\mathbf{b}, \mathbf{c}\}$, and $\{\mathbf{b}, \mathbf{d}\}$, because they yield knowledge state $\varnothing$. The three remaining sets of two skills give the states $\{1\}$, $\{2\}$, and $\{3\}$, which are also generated by some sets of three skills. For example, $\{\mathbf{a}, \mathbf{b}, \mathbf{d}\}$ yields $\{1\}$. The states $\{1, 3\}$, $\{2, 3\}$ are generated by the two sets of three skills $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ and $\{\mathbf{a}, \mathbf{c}, \mathbf{d}\}$, respectively. However, no individual can be in State $\{1, 2\} = \{1\} \cup \{2\}$ because any subject having mastered Problems 1 and 2 necessarily has the skills **a**, **b**, **c**, and **d** and is thus in State $\{1, 2, 3\}$. These examples show that there may be some states in a knowledge space as we define it that cannot be realized in practice. Our position with respect to such examples is as follows.

For the purpose of this discussion, we refer to any state of a knowledge space which can never be realized empirically as *fictitious*. We do not deny that there may be situations in which a knowledge space contains fictitious states. Our working hypothesis is that the number of such states will be relatively small. Because our main goal is to have a knowledge structure in which all subjects in a population of interest can be described, the fact that there may be a few descriptions that will never be used is of little consequence. This strategy thus involves enlarging the collection of states for the sake of simplicity. In that light, the $\cup$-closure axiom is in line with many idealizing assumptions in scientific theory. However, if the number of fictitious states is large and if considering such states renders the knowledge assessment too costly, then this assumption will have to be dropped, at least at the ultimate stage of our construction of the knowledge space. However, notice that it can still play a useful role in the intermediate stage. We recall that the determination of the knowledge space was accomplished in two major stages. In the Building the Knowledge Space section, we showed how a first sketch of the knowledge space could be obtained by systematic questioning of experts, using the algorithm QUERY. Successive steps of this algorithm yield smaller and smaller knowledge spaces. In the second phase, which is based on the analysis of empirical data in terms of the stochastic learning theory, the remaining set of states is further refined. At that point, the assumption that the knowledge structure is $\cup$-closed can be abandoned. We could, for instance, simply suppose that the knowledge structure satisfies the axiom of well-gradedness (cf. Definition 4). No major damage to the theory would ensue.

Another objection is that no explicit distinction is made in our work between procedural knowledge and factual knowledge. Our contention is that such a distinction, however interesting it may be for other purposes, is not essential in our approach. A procedure can practically always be cast as a problem (e.g., "Give a proof by induction of the following fact").

More generally, some critics argue that the description of the knowledge space, as the term is understood in our work, is not very detailed. Consider the case of a student failing to solve a multiplication problem. One may wish to know the type of error made and have the mechanism of the error included in the description of the student's knowledge state. However, this objection overlooks the obvious possibility of having a wide and highly structured array of multiplication problems, the responses to which would provide, within the same knowledge space framework, a precise diagnostic of the faulty algorithm used. (One might have one-digit multiplications, two-digit multiplications, multiplications involving decimals, etc.) We venture that any error mechanism can always be specified, as accurately as one might wish, by the collections of problems solved and failed.

An important set of results at the focus of our work concerns the mathematical equivalence between three distinct concepts: the knowledge spaces, the surmise systems, and the relation P at the basis of the QUERY procedure for building the knowledge space. The additional condition of well-gradedness for knowledge spaces was introduced at a later stage and was found to play an important role, both in the sections on the description and application of the stochastic learning theory and in the previous section on the discrete Markov procedure. It is natural to ask what the corresponding condition is for surmise systems and relation P. We have only a partial solution to this question; we know what condition on a surmise system renders this concept equivalent to a well-graded knowledge space (see Koppen, 1988). However, we do not know what particular kind of relation P corresponds exactly to a well-graded knowledge space. This is a weakness, in that, for the time being, we cannot adapt the QUERY procedure for constructing a well-graded knowledge space. We are hopeful that this difficulty will be solved in the near future.

Turning to the learning theory, we recall several minor, somewhat technical drawbacks. Some of our assumptions were made for convenience and had no other justification than the need to keep our model within manageable bounds. One example is the supposition that the learning rate remains constant in the course of learning. Another concerns the assumed independence between the choice of a gradation and the learning rate. These assumptions are certainly not realistic and would certainly fail under close examination. The distributional assumptions on the learning rate and on the learning latencies may also appear rather arbitrary.

Standard arguments can be given to defend such assumptions. Mathematical simplicity, for the sake of rendering a model applicable, should not be hastily and automatically discarded as an unworthy goal because it is conceivable, and one can always hope, that the unrealistic assumptions actually make little difference in the predictions. This hope was justified by the analysis of the data. Not only did we obtain a good fit of the model to realistic data (cf. Table 5 and Figure 7), but a much more specific validation was also recorded. Our main interest in the stochastic learning theory was in the convenient statistical analysis of the data that it provided, which could lead to successive simplification of the well-graded knowledge space. In other words, we were not primarily interested in the details of the theory. Rather, we wanted to use it to confirm and refine a particular knowledge space to pave the way for an application of the knowledge assessment procedures. In this respect, side assumptions such as those concerning the distributions were just useful technical devices. This position received at least partial support in the statistical results obtained from drastically modifying the distributional assumptions. This was discussed in the learning theory application section, in which we reported that an equally good fit to the four-gradations model was obtained when we changed the expressions of the incomplete beta function ratio in Equations 20 and 21 to express a completely different distribution function, defined by the composition of the logit function by the standard normal distribution function.

A large-scale application of these techniques is in progress, based on a knowledge space with 50 problems. We also plan to undertake a systematic comparison between human experts examining students and computerized assessment procedures of the type described in the previous section.

The theoretical constructions of this article depart from traditional approaches to psychometric testing. As recalled in the Knowledge Spaces and Skills section, psychometric models are typically based on the concept of skills or abilities. Most often, these concepts are given a numerical interpretation in the framework of a model that is completely specified except for the values of some parameters. There is no doubt that knowledge can be assessed by way of such models (see, for instance, Weiss, 1983, for articles reviewing this line of work). The knowledge state of an individual can be identified with some ability level, which can be uni- or multidimensional. The assessment proceeds by successively asking several questions to the individual and gradually refining some estimates of parameters in the model, representing ability levels. These numbers can then be used to predict actual knowledge. Whether psychometric approaches to knowledge assessment are more efficient than those described in this article is left for further investigation. We must, however, emphasize a fundamental difference between standard psychometric models and the theoretical constructions discussed in this article.

The standard psychometric models can be rightly described as quantitative, with the usual meaning of this term in science, which has a strong numerical connotation. These psychometric models, in which intellectual competence is assessed by a small number of abilities evaluated numerically, are simple and convenient representations. It must be understood that, until very recently, numerical models were essential because they were the only models capable of being analyzed in any depth, for which detailed predictions could actually be worked out. Simplicity could have been operationally defined by the fact that the analysis of the model could be performed with a desk calculator or a first-generation computer.

The situation has changed dramatically in the last two decades or so, with the coming of age of much more powerful computers. The models now developed by scientists are often profoundly different from what they were. The concept of simplicity (as applied to scientific theories) itself may be changing. As a consequence, more realistic models, combinatorial in character, may be considered. The work presented in this article offers an example of this trend. Twenty-five years ago, such modeling would have made little sense and could have been discarded as mere speculation without any practical use.

References

Birkhoff, G. (1937). Rings of sets. *Duke Mathematical Journal, 3,* 443–454.

Brent, R. P. (1973). *Algorithms for minimization without derivatives.* Englewood Cliffs, NJ: Prentice-Hall.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Degreef, E., Doignon, J.-P., Ducamp, A., & Falmagne, J.-C. (1986). Languages for the assessment of knowledge. *Journal of Mathematical Psychology, 30,* 243–256.

Doignon, J.-P., & Falmagne, J.-C. (1985). Spaces for the assessment of knowledge. *International Journal of Man–Machine Studies, 23,* 175–196.

Doignon, J.-P., & Falmagne, J.-C. (1987). Knowledge assessment: A set theoretical framework. In B. Ganter, R. Wille, & K. E. Wolff (Eds.), *Beiträge zur Begriffsanalyse, Vorträge der Arbeitstagung Begriffsanalyse, Darmstadt, 1986* (pp. 129–140). Mannheim, Federal Republic of Germany: B. I. Wissenschaftsverlag.

Doignon, J.-P., & Falmagne, J.-C. (1988). Parametrization of knowledge structures. *Discrete Applied Mathematics, 21,* 87–100.

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis.* New York: Wiley.

Falmagne, J.-C. (1989a). A latent trait theory via a stochastic learning theory for a knowledge space. *Psychometrika, 54,* 283–303.

Falmagne, J.-C. (1989b). Probabilistic knowledge spaces: A review. In F. Roberts (Ed.), *Applications of combinatorics and graph theory to the biological and social sciences* (IMA Vol. 17). New York: Springer-Verlag.

Falmagne, J.-C., & Doignon, J.-P. (1988a). A class of stochastic procedures for the assessment of knowledge. *British Journal of Mathematical and Statistical Psychology, 41,* 1–23.

Falmagne, J.-C., & Doignon, J.-P. (1988b). A Markovian procedure for assessing the states of a system. *Journal of Mathematical Psychology, 32,* 232–258.

Feller, W. (1970). *An introduction to probability and its applications* (3rd ed., rev.). New York: Academic Press.

Fu, K. S. (1974). *Syntactic methods in pattern recognition.* New York: Academic Press.

Johnson, N. L., & Kotz, S. (1970). *Distributions in statistics: Continuous univariate distributions* (Vol. 2). New York: Houghton Mifflin.

Kemeny, J. G., & Snell, J. L. (1965) *Finite Markov chains.* Princeton, NJ: Van Nostrand.

Koppen, M. (1988). *Surmise mappings and well-graded knowledge spaces.* Manuscript submitted for publication.

Koppen, M., & Doignon, J. P. (in press). How to build a knowledge space by querying an expert. *Journal of Mathematical Psychology.*.

Landy, M. S., & Hummel, R. A. (1986). A brief survey of knowledge aggregation methods. In *Proceedings of the International Conference on Pattern Recognition* (pp. 248–252).

Lazarsfeld, P. F. (1959). Latent structure analysis. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 3). New York: McGraw-Hill.

Lord, F. M. (1974). Individualized testing and item characteristic curve theory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: Vol. 2. Measurement, psychophysics, and neural information processing.* San Francisco: Freeman.

Lord, F. M., & Novick, M. R. (1974). *Statistical theories of mental tests* (2nd ed.). Reading, MA: Addison-Wesley.

Marshall, S. P. (1981). Sequential item selection: Optimal and heuristic policies. *Journal of Mathematical Psychology, 23,* 134–152.

Monjardet, M. (1970). Tresses, fuseaux, préordres et topologies [This paper deals with braids, quasi-order, and topologies]. *Mathématiques et Sciences Humaines, 30,* 11–22.

Powell, M. J. D. (1964). An efficient method for finding the minimum of a function in several variables without calculating derivatives. *Computer Journal, 7,* 155–162.

Roberts, F. S. (1979). Measurement theory with applications to decision making, utility and the social sciences. In Gian-Carlo Rota (Ed.), *Encyclopedia of mathematics and its applications: Vol. 7. Mathematics and the social sciences.* Reading, MA: Addison-Wesley.

Shortliffe, E. H. (1976). *Computer-based medical consultation: Mycin.* New York: American Elsevier.

Shortliffe, E. H., & Buchanan, D. G. (1975). A model of inexact reasoning in medicine. *Mathematical Bioscience, 23,* 351–379.

Villano, M., Falmagne, J.-C., Johannesen, L., & Doignon, J.-P. (1987). Stochastic procedures for assessing an individual's state of knowledge. In *Proceedings of the International Conference on Computer-Assisted Learning in Post-Secondary Education* (pp. 369–371). Calgary, Alberta, Canada: University of Calgary.

Wainer, W., & Messick, S. (1983). *Principles of modern psychological measurement: A festschrift for Frederic M. Lord.* Hillsdale, NJ: Erlbaum.

Weiss, D. J. (Ed.). (1983). *New horizons in testing: Latent trait theory and computerized testing.* New York: Academic Press.

Appendix

The *gamma function* is a real valued function Γ, defined for all positive real numbers α by the integral

$$\Gamma(\alpha) = \int_0^\infty \lambda^{\alpha-1} e^{-\lambda} d\lambda. \qquad (A1)$$

64

The *beta function* $(p, q) \rightarrow B(p, q)$ is defined by the equation

$$B(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p + q). \qquad (A2)$$

65

For $\alpha, \xi > 0$, we derive from Equation A1

$$[\xi^\alpha/\Gamma(\alpha)] \int_0^\infty \lambda^{\alpha-1} e^{-\xi\lambda} d\lambda = 1.$$

66

This result leads to defining the density f of a *general gamma* random variable with parameters $\alpha, \xi > 0$, by the formula

$$f(\lambda) = [\xi^\alpha/\Gamma(\alpha)]\lambda^{\alpha-1}e^{-\xi\lambda}, \text{ for any } \lambda > 0,$$

67

with $f(\lambda) = 0$ for $\lambda \leq 0$. The incomplete beta function ratio of Equation 19 is defined in terms of the beta function specified by Equation A2.

Received July 5, 1988
Revision received March 11, 1989
Accepted April 25, 1989 ■