PROBABILITY-BASED INFERENCE IN A DOMAIN OF PROPORTIONAL REASONING TASKS

Anne Béland
Université de Sherbrooke

Robert J. Mislevy
Educational Testing Service

This research was sponsored in part by the
Cognitive Science Program
Cognitive and Neural Sciences Division
Office of Naval Research, under
Contract No. N00014-88-K-0304
R&T 4421573

Robert J. Mislevy, Principal Investigator

Educational Testing Service
Princeton, New Jersey

January 1992

Reproduction in whole or in part is permitted
for any purpose of the United States Government.

Approved for public release; distribution unlimited.
Educators and psychologists are increasingly interested in modelling the processes and knowledge structures by which people learn and solve problems. Progress has been made in developing cognitive models in several domains, and in devising observational settings that provided clues about subjects' cognition from this perspective. Less attention has been paid to procedures for inference or decision-making with such information, given that it provides only imperfect information about cognition - in short, test theory for cognitive assessment. This paper describes probability-based inference in this context, and illustrates its application with an example concerning proportional reasoning.

Key words: Bayesian inference, cognitive assessment, inference networks, multiple strategies, proportional reasoning, test theory.
Probability-Based Inference in a Domain of Proportional Reasoning Tasks

Anne Béland
Université de Sherbrooke

Robert J. Mislevy
Educational Testing Service

January, 1992

Authors' names appear in alphabetical order. Work upon which this paper is based was carried out under Dr. Béland's postdoctoral fellowship at Educational Testing Service. Dr. Mislevy's work was supported by Contract No. N00014-91-J-4101, R&T 4421573-01, from the Cognitive Science Program, Cognitive and Neural Sciences Division, Office of Naval Research, and by ETS's Program Research Planning Council. We are grateful for Duanli Yan's technical assistance in implementing the inference network, and for Duanli's and Kathy Sheehan's comments on an early draft of the paper.
Probability-Based Inference in a Domain of Proportional Reasoning Tasks

Abstract

Educators and psychologists are increasingly interested in modelling the processes and knowledge structures by which people learn and solve problems. Progress has been made in developing cognitive models in several domains, and in devising observational settings that provide clues about subjects' cognition from this perspective. Less attention has been paid to procedures for inference or decision-making with such information, given that it provides only imperfect information about cognition—in short, test theory for cognitive assessment. This paper describes probability-based inference in this context, and illustrates its application with an example concerning proportional reasoning.

Key words: Bayesian inference, cognitive assessment, inference networks, multiple strategies, proportional reasoning, test theory
Introduction

The view of human learning rapidly emerging from cognitive and educational psychology emphasizes the active, constructive role of the learner in acquiring knowledge. Learners become more competent not simply by learning more facts and skills, but by configuring and reconfiguring their knowledge; by automating procedures and chunking information to reduce memory loads; and by developing models and strategies that help them discern when and how facts and skills are relevant. Educators have begun to view school learning from this perspective, as a foundation for instruction in both the classroom and intelligent computer-assisted instruction, or intelligent tutoring systems (ITSs).

Making educational decisions cast in this framework requires information about students in the same terms. Glaser, Lesgold, and Lajoie state,

Achievement testing as we have defined it is a method of indexing stages of competence through indicators of the level of development of knowledge, skill, and cognitive process. These indicators display stages of performance that have been attained and on which further learning can proceed. They also show forms of error and misconceptions in knowledge that result in inefficient and incomplete knowledge and skill, and that need instructional attention. (Glaser, Lesgold, & Lajoie, 1987, 81)

Standard test theory is designed to characterize students in terms of their tendencies to make correct answers, not in terms of their skills, strategies, and knowledge structures. Yet generalizations of the questions that led to standard test theory arise immediately in the context Glaser and his colleagues describe: How can we design efficient observational settings to gather the data we need? How can we make and justify decisions? How do we evaluate and improve the quality of our efforts? Without a conceptual framework for inference, rigorous answers to these questions are not forthcoming.

This presentation addresses issues in model building and statistical inference in the context of student modelling. The statistical framework is that of inference networks (e.g.,
Probability-Based Inference in Cognitive Assessment

Comparing the ways experts and novices solve problems in domains such as physics and chess (e.g., Chi, Feltovich & Glaser, 1981) reveals the central importance of knowledge structures—interconnected networks of concepts referred to as "frames" (Minsky, 1975) or "schemas" (Rumelhart, 1980)—that impart meaning to observations and actions. The process of learning is, to a large degree, expanding these structures and, importantly, reconfiguring them to incorporate new and qualitatively different connections as the level of understanding deepens. Researchers in science and mathematics education have focused on identifying key concepts and schemas in these content areas, studying how they are typically acquired (e.g., in mechanics, Clement, 1982; in proportional reasoning, Karplus, Pulos, & Stage, 1983), and constructing observational settings in which students' understandings can be inferred (e.g., van den Heuvel, 1990; McDermott, 1984). A key feature of most of these studies is explaining patterns observed in learners' problem-solving behavior in terms of their knowledge structures. Riley, Greeno, and Heller (1983), for example, explain typical patterns of errors and correct answers in children's word problems in terms of a hierarchy of successively sophisticated procedural models.

Once the relevance of states of understanding to instructional decisions is accepted, one immediately confronts the fact that these states cannot be ascertained with certainty;
they can be inferred only imperfectly from observations of the students' behavior. Research in subject areas is beginning to provide observational situations (at their simplest form, test items) that tap particular aspects of knowledge structures (e.g., Lesh, Landau, & Hamilton, 1983; Marshall, 1989). Conformable statistical models must be capable of expressing the nature and the strength of evidence that observations convey about knowledge structures. Two kinds of variables are thus involved: those expressing characteristics of an inherently unobservable student model, and those concerning qualities of observable student behavior, the latter of which presumably carry information about the former.

For the special case in which a student is adequately characterized by a single unobservable proficiency variable, a suitable statistical methodology has been developed within the paradigm of standard test theory, most notably under the rubric of item response theory (IRT; see Hambleton, 1989). IRT posits a model for the probability of a correct response to a given test item, as a function of parameters for the examinee's proficiency (often denoted $\theta$) and measurement properties of the item. The IRT model provides the structure through which observable responses to test items are related to one another and to the unobservable proficiency variables. Item parameters specify the degree or strength of relationships within that structure, by quantifying the conditional probabilities of item responses given $\theta$. Observed item responses induce a likelihood function for $\theta$, opening the door to statistical inference and decision-making models. The coupling of probability-based inference with a simple student model for overall proficiency provides the foundation for item development, test construction, adaptive testing, test equating, and validity research—all providing, of course, that "overall proficiency" is sufficient for the job at hand.

Models connecting observations with a broader array of cognitively-motivated unobservable variables have begun to appear in the psychometric literature. Table 1 offers
a sampling. The approach we have begun to follow continues in the same spirit. In any given implementation, the character of unobservable variables and the structure of their interrelationships is derived from the structure and the psychology of the substantive area, with the goal of capturing key distinctions among students. Probability distributions characterize the likelihoods of potential observable variables, given values of the variables in the unobservable student model. The relationship of the observable variables to the unobservable variables characterizes the nature and amount of information they carry.

[Insert Table 1 about here]

Of particular importance is the concept of conditional independence: a set of variables may be interrelated in a population, but independent given the values of another set of variables. In cognitive models, relationships among observed variables are "explained" by inherently unobservable, or latent, variables. Pearl (1988) argues that creating such intervening variables is not merely a technical convenience, but a natural element in human reasoning:

"...conditional independence is not a grace of nature for which we must wait passively, but rather a psychological necessity which we satisfy actively by organizing our knowledge in a specific way. An important tool in such organization is the identification of intermediate variables that induce conditional independence among observables; if such variables are not in our vocabulary, we create them. In medical diagnosis, for instance, when some symptoms directly influence one another, the medical profession invents a name for that interaction (e.g., 'syndrome,' 'complication,' 'pathological state') and treats it as a new auxiliary variable that induces conditional independence; dependency between any two interacting systems is fully attributed to the dependencies of each on the auxiliary variable."  
(Pearl, 1988, p. 44)
Inference Networks

A heritage of statistical inference under the paradigm described above extends back beyond IRT, to Charles Spearman's (e.g., 1907) early work with latent variables, Sewell Wright's (1934) path analysis, and Paul Lazarsfeld's (1950) latent class models. The resemblance of the inference networks presented below to LISREL diagrams (Jöreskog & Sörbom, 1989) is no accident! The inferential logic of test theory is built around conditional probability relationships—specifically, probabilities of observable variables given theoretically-motivated unobservable variables.

The starting point is a recursive representation of the joint distribution of a set of random variables; that is,

\[ p(X_1, \ldots, X_n) = p(X_n|X_{n-1}, \ldots, X_1) \cdot p(X_{n-1}|X_{n-2}, \ldots, X_1) \cdots p(X_2|X_1) \cdot p(X_1) \]

\[ = \prod_{j=1}^{n} p(X_j|X_{j-1}, \ldots, X_1) , \]

(1)

where the term for \( j=1 \) is defined as simply \( p(X_1) \). A recursive representation can be written for any ordering of the variables, but one that exploits conditional independence relationships can be more useful. For example, under an IRT model with one latent proficiency variable \( \theta \) and three test items, \( X_1, X_2, \) and \( X_3 \), it is equally valid to write

\[ p(X_1, X_2, X_3, \theta) = p(\theta|X_3, X_2, X_1) \cdot p(X_3|X_2, X_1) \cdot p(X_2|X_1) \cdot p(X_1) \]

(2)
or

\[ p(X_1, X_2, X_3, \theta) = p(X_3|X_2, X_1, \theta) \cdot p(X_2|X_1, \theta) \cdot p(X_1|\theta) \cdot p(\theta) . \]

(3)

But (3) simplifies to

\[ p(X_1, X_2, X_3, \theta) = p(X_3|\theta) \cdot p(X_2|\theta) \cdot p(X_1|\theta) \cdot p(\theta) . \]

(4)
the form that harnesses the power of IRT by expressing test performance as the
concatenation of conditionally independent item performances. More generally, (1) can be
re-written as
\[ p(X_1, \ldots, X_n) = \prod_{j=1}^{n} p(X_j|\text{parents of } X_j), \]

where \( \text{parents of } X_j \) is the subset of variables upon which \( X_j \) is directly dependent.

Corresponding to the algebraic representation of \( p(X_1, \ldots, X_n) \) in (5) is a graphical
representation—a directed acyclic graph (DAG). Each variable is a node in the graph;
directed arrows run from parents to children, indicating conditional dependence
relationships among the variables. In this paper we refer to such a structure or its graphical
representation as an inference network. Figure 1 shows the DAGs that correspond to (2)
and (4) in the IRT example. Note that the simplified structure is apparent only in the graph
for (4). A DAG does not generally reveal conditional independence relationships that might
arise under alternative orderings of the variables.

Different fields of application emphasize different aspects of inference network
representations of systems of variables. In factor analyses of mental tests, for example,
one important objective is to find a “simple structure” representation of the relationships
among test scores, wherein each test has only a few latent variables as parents (e.g.,
Thurstone, 1947). In sociological and economic applications, path analysis is used to sort
out the direct and indirect effects of selected variables upon others (e.g., Blalock, 1971).
In animal husbandry, where genotypes are latent nodes and inherited characteristics of
animals are observable, interest lies in the predicted distribution of characteristics of the
offspring of potential matings (e.g., Hilden, 1970). In medical diagnosis, disease states
and syndromes are unobserved nodes, while symptoms and test results are potential
observables; ascertaining the latter guides diagnosis and treatment decisions (e.g., Andreassen, Jensen, & Olesen, 1990).

The latter arenas have sparked interest in calculating distributions of remaining variables conditional on observed values of a subset. If the topology of the DAG is favorable, such calculations can be carried out in real time in large systems by means of local operations on small subsets of interrelated variables ("cliques") and their intersections. The interested reader is referred to Lauritzen and Spiegelhalter (1988), Pearl (1988), and Shafer and Shenoy (1988) for updating strategies, a kind of generalization of Bayes theorem. The calculations for the following example were carried out with Andersen, Jensen, Olesen, and Jensen's (1989) HUGIN computer program.

The point of this presentation is that inference networks can be constructed around cognitive student models. The analogy to medical applications is sketched in Table 2. A key aspect of the correspondence is the flow of diagnostic reasoning: Theory is expressed in terms of conditional probabilities of observations given theoretically suggested unobservable variables, and it is from this direction that the inference network is constructed. Reasoning in practical applications flows in the opposite direction, as evidence from observations is absorbed, to update belief about the unobservable variables. This necessity of bidirectional reasoning stimulates interest in probability-based inference, as accomplished by the generalizations of Bayes Theorem mentioned above.

[Insert Table 2 about here]

An Inference Network for a Set of Juice-Mixing Tasks

Proportional reasoning is a topic of great current interest among mathematics and science educators, because it constitutes perhaps half of the middle school mathematics curriculum, and is a prerequisite for quantitative aspects of the sciences as well as advanced topics in mathematics. There is consequently considerable research on this topic among the communities of both
developmental psychology (e.g., Inhelder & Piaget, 1958; Siegler, 1978) and the psychology of
mathematics education (e.g., Romberg, Lamon, & Zarinnia, 1988). The network presented here is
based on a program of research on the development of proportional reasoning represented by
Noelting (1980a; 1980b) and Béland (1990). According to this conceptual framework, subjects’
cognitive strategies are explained in terms of the relationships they address vis a vis the structural
properties of the items. Development is viewed as a progression through qualitatively distinct
levels of understanding.

In order to study the concept of proportion, a basic test of twenty items was
devised. Each consisted of predicting the relative taste of two drinks, labeled A and B,
which comprised varying numbers of glasses of juice and glasses of water. Each mixture
defined an ordered pair, that is \((a, b)\) for the drink labeled A, and \((c, d)\) for the drink
labeled B. The first term in each pair defined the number of glasses of juice and the second
term defined the number of glasses of water, as shown by the example in Figure 2. In the
test, the child had to decide if either A or B would taste juicier, or if both drinks would taste
the same. The subjects also had to explain the reasons for their choices by writing a
detailed explanation of how they had solved each problem. A total number of 448 subjects,
ranging from fourth graders to university freshman, were assessed. Instructions were
given and data collected in class groups. The order of item presentation was randomized
for each child. To assure that the task was understood, sample items were solved by the
classes.

An item’s components were differentiated as being the varying quantities of juice glasses,
which defined the attribute, and water glasses, which defined the complement, in each pair. When
a subject attempted to solve an item by constructing transformations between similar terms in both
pairs, that is, either between the attribute or the complement in both mixtures, then the relationships
were described as scalar. On the other hand, when the transformations were constructed between
the complementary terms within each pair, that is, between the attribute and the complement in a mixture, then the relationships were described as functional. Three qualitatively distinct ordered levels (listed below) were defined as a set of additive and multiplicative relations among the values of these terms. These levels characterize both items and solution strategies: solution strategies, in terms of the kinds of transformations and comparisons they involve; items, by virtue of their structure, in terms of the minimal level required for a correct understanding of the problem. The fact that some strategies led to success with items at one level, but to failure with items at higher levels, indicates a structural discontinuity between these levels. This implies that the transition between these levels involves restructuring, or reconceptualizing, the relationships among task components, in response to the structural properties of the items. The three levels of understanding are as follows.

- **Level 1**, the *preoperational* level, is characterized by the differentiation and coordination of scalar and functional relationships. For example, one justification for solving the item (2,1) vs. (3,4) was: "Mixture A tastes juicier because the number of juice glasses is greater than the number of water glasses. By comparison, mixture B tastes less juicy because the quantity of water glasses is greater than juice glasses."

- **Level 2**, the *concrete operational* level, is characterized by the construction of an equivalence class. For example, to solve the item (2,6) vs. (3,9), the typical justification for the functional operator was: "Both drinks taste alike because there is one glass of juice for three glasses of water, which defines the ratio 1:3 in both pairs."

- **Level 3**, the *formal operational* level, is characterized by the construction of a combinatorial system, building upon the concepts from the previous levels. An item is solved either by the *between* state ratios (common denominator) or the *within* state ratios (percentage). For example, when a ratio strategy was used to
solve (3,5) vs. (2,3), the typical justification was: “In Mixture A there are three
glasses of juice for five glasses of water, a ratio of 9:15. In Mixture B the ratio is
10:15 juice to water. Therefore, B tastes juicier.”

The gradual extension of these structures, through exercise and practice, leads to the
consolidation of the cognitive strategies as they are applied to solve the increasing complexity of
the items within a level. This progression was defined as stage within level. Three successive
stages, denoted as a, b, and c, were defined within each level. Table 3 summarizes the stages
within levels. The reader is referred to Béland (1990) for additional detail and discussion.

[Insert Table 3 about here]

An Overview of the Network

An inference network was constructed on the basis of the data described above,
addressing subjects’ optimal cognitive stage x level, or the highest stage and level at which
they were observed to perform during the course of observation, and the details of their
responses to three items, one at each level. This section introduces the network. The
following section describes the variables in more detail, and discusses the specification of
conditional probabilities. The section after that gives examples of reasoning from
observations back to cognitive levels.

The network addresses the three items shown in Figure 3, which appeared as 3, 8,
and 17 in the master list. Item 3, (2,1) vs. (3,4), is a level 1 item, since it can be correctly
solved by a level 1 strategy: Mixture A has more juice than water, while B has more water
than juice. Item 8, (2,6) vs. (3,9), is a level 2 item, since it requires the construction of an
equivalence class. Item 17, (3,5) vs. (2,3), is a level 3 item, since a solution that correctly
attends to its structure must, in some way, compare ratios.

[Insert Figure 3 about here]
The 21 variables in the network are listed below, with the number of possible values each variable can take in parentheses. Detailed descriptions appear in the following section.

- $X_1$: Optimal cognitive level (3).
- $X_2$: Stage within optimal level (3).
- $X_3$: Optimal stage x level (9).
- $X_{4j}$: Strategy employed on Item j, for j=3, 8, and 17 (10 per item).
- $X_{5j}$: Procedural analysis for Item j (4 per item).
- $X_{6j}$: Understanding of structure of Item j (2 per item).
- $X_{7j}$: Solution of Item j (2 per item).
- $X_{8j}$: Response choice on Item j (3 per item).
- $X_{9j}$: Objective correctness of response choice on Item j (2 per item).

Without constraints, the joint distribution of the variables listed above would be specified as a probability for each of the $3 \times 3 \times 9 \times (10 \times 4 \times 3 \times 2 \times 2 \times 2)^3$ possible combinations of values—about $7 \times 10^{10}$ of them. Under the assumed network, however,

$$p(X_1, X_2, X_3, X_{4,3}, X_{4,4}, X_{4,8}, X_{4,17}, ..., X_{9,3}, X_{9,8}, X_{9,17})$$

$$= p(X_1) p(X_2|X_1) p(X_3|X_2, X_1)$$

$$\times \prod_j p(X_{4j}|X_3) p(X_{5j}|X_{4j}) p(X_{6j}|X_{5j}) p(X_{7j}|X_{5j}) p(X_{8j}|X_{5j}, X_{4j}) p(X_{9j}|X_{8j}).$$

(6)

As examples, (6) implies conditional independence of item responses, $X_{4,3}, X_{4,8},$ and $X_{4,17},$ given a subject's optimal cognitive stage x level, $X_3$ (although we discuss below relaxing this assumption to account for processes that characterize the adaptive quality of children's strategy choices during the course of testing); and conditional independence of the correctness of the response choice for Item j, $X_{9j},$ from all other variables given the identity of that response choice, $X_{8j}$. The most complex of these local relationships in (6) involves only three variables, and the total number of distinct probabilities needed to approximate the full joint distribution is $3+9+81+
3(90+40+120+8+8+6), or 909. As we shall see, many of these relationships are logical rather than empirical, and can be specified without recourse to data.

Figure 4 is the DAG corresponding to (6). Figure 5 is a similar graph from HUGIN, exhibiting for each node the baseline marginal distribution for each variable with bars representing the probabilities for each potential value of a variable. These population base rates were established from the responses of all subjects, as described in the next section. Figure 5 represents the state of knowledge one would have as a new subject from the same population is introduced. As she makes responses, the relevant nodes will be updated to reflect certain knowledge of, say, the correctness of a response or the strategy level used to justify it. This would be represented by a probability bar extending all the way to one for the observed value. This information updates (still imperfect) knowledge about her optimal cognitive level, and expectations about what might be observed on subsequent items.

[Insert Figures 4 and 5 about here]

Instantiating the Network

The initial status of the network is the joint distribution of all the variables. It is specified via (6) in terms of the baseline distribution of any variables without parents, and the conditional distributions of each of the remaining variables given its parents. Béland’s classifications of all response explanations of all subjects into stage x level categories were employed, and treated as known with certainty.¹ Explanations of the variables and discussions of the conditional probabilities associated with each follow.

¹ A small proportion of the response strategies could not be classified, because subjects’ explanations were either omitted or incomprehensible. These responses were not useful in determining a subject’s highest strategy level, but they were included in the following analyses, with “undifferentiated” as a potential value of strategy choice. The proportions for Items 3, 8, and 17 were 2%, 1%, and 11% respectively.
**X1: Optimal cognitive level.** Each subject was classified as to the stage and level of his or her highest level solution strategy, based on Béland's analyses of all twenty of their response explanations. $X_1$ denotes their highest level, collapsing over stages within levels. Because it has no parents, we need specify only population proportions: .08 for Level 1, .45 for Level 2, and .47 for Level 3.

**X2: Stage within optimal level.** $X_2$ breaks down stage membership within levels, so $X_1$ is its parent. Empirical proportions were employed, leading to the values shown in Table 4. Again these values are based on Béland's classification. Among the subjects whose highest observed level of solution strategy was Level 2, for example, what proportions of these highest strategies were at Stages a, b, and c of Level 2? Stages are meaningful only within levels, so the marginal distribution of $X_2$ that appears in Figure 5 is not very useful. If $X_1$ were fixed at a particular value of level, however, the resulting marginal distribution for $X_2$ would be meaningful, taking the values from the appropriate row of Table 4.

[Insert Table 4 about here]

**X3: Optimal stage x level.** $X_3$ is the detailed categorization of subjects into mutually exclusive and exhaustive categories, in terms of levels and stages. It has as parents both level, $X_1$, and stage within level, $X_2$. The specification of conditional probabilities under this arrangement is logical rather than empirical: The conditional probability of a given stage-within-level value is 1 only if $X_1$ and $X_2$ take the appropriate values; otherwise, the conditional probability is zero. This can be seen in Figure 6, where conditioning on $X_3=3b$ leads to probabilities of one for Level=3 and Stage-within-level=b. Actually no information would be lost by having $X_1$ and $X_2$ but not $X_3$ in the model, or $X_3$ but not $X_1$ and $X_2$. We have included all of them for interpretive convenience; for example, $X_1$ is useful for summarizing the "level" information in $X_3$, whereas the values for $X_3$ lie at the same level of detail as those of the Item Strategy variables described below.
Under the "dialectical constructivist" developmental model sketched above, a subject's optimal structure level defines the repertoire of strategies available for solving a given item, as constructed through the changes and transformations that the subjects generated during the course of testing. That is, the optimal state of understanding was constructed by the learners through a series of mental operations that defined the successive levels of conceptualization elaborated to seek the structural properties of the item. Consequently, the optimal structure was not necessarily operationalized before the subjects undertook the task. The dynamics of this process are not modelled in the present example, but will be discussed below. Conversely, the strategy required to solve a given problem was not ultimately at the same level as the subject's optimal stage x level, even when that level has been attained. This observation is taken into account in the present model, through the conditional probability matrices for the following item strategy variables.

\[ X_{4j}; \text{Strategy employed on Item } j \ (j=3, 8, 17). \] In addition to subjects' optimal strategy stage x level, the particular strategies they employed in the three exemplar items were classified according to stage x level, constituting the variables \( X_{4j} \). The additional value, abbreviated "Ud" in the HUGIN diagrams, stands for "Undifferentiated;" these are the responses which could not be classified. The \( X_{4j} \) variables are modelled as conditionally independent, given their common parent \( X_3 \), optimal cognitive level. The conditional probability matrices are presented in Table 5.

The following features are noting:

- With a few exceptions, a strategy at any level could be applied to any item. A small number of "logical zeros" appear when the conceptual elements in a given strategy class had no possible correspondents in the structure of an item (e.g., a 2b strategy for Item 17).
• The entire upper right triangle of each matrix is filled with "logical zeros." By definition, it is not possible to observe a response strategy at a higher stage x level than a subject's optimal stage x level.

• The lower left triangle of each matrix was estimated empirically for the most part, by simply entering the proportion of subjects classified in a given optimal stage x level who were classified as employing each of the response strategies for a given item. Probabilities that were logically possible but empirically zero were replaced by small positive probabilities. It can be seen that considerable variation in strategy choice on a given item often existed among subjects with the same optimal level. Among subjects whose optimal stage x level was 3b, for example, about half employed this powerful strategy for the more simply structured Item 8, while about 40% adapted their strategies to the structure of the item and employed a "minimally sufficient" strategy at level 2b. This information appears graphically in Figure 6.

[Insert Table 5 about here]

X_{5j}: Procedural analysis for Item j. These variables summarize the results of the matchups between cognitive strategies and qualitative outcomes. The four possible values are "Success," in which a strategy at the same level as (isomorphic to) the item, or higher, was successfully employed; "Strategic error," in which a strategy was employed which failed to account for the item's structure; "Tactical error," in which a strategy appropriate to the item structure was employed but not successfully executed; and "Computational error," in which the attempt would have been a "Success" except for an error in numerical calculations. The respective X_{4j} variables are the parents. Conditional probabilities corresponding to "Strategic error" are logical, since this outcome is necessary if a strategy that is insufficient vis a vis the item structure is applied, and impossible if a sufficient
strategy is applied.\(^2\) In the latter case, conditional probabilities are apportioned among “Success,” “Tactical error,” and “Computational error.” Table 6 lists the conditional probability values.

[Insert Table 6 about here]

\(X_{5j} \): **Understanding of structure of Item j.** These variables simply collapse from their parents, the \(X_{5js}\), into the dichotomy of “Understood” or “Misunderstood” the structural properties of the item. In each case, the conditional probability matrix is logical: the probability for “Understood” is one if the procedural analysis is “Success,” “Tactical error,” or “Computational error,” and zero otherwise; the probability for “Misunderstood” is one if the procedural analysis is “Strategic error,” and zero otherwise.

\(X_{7j} \): **Solution of Item j.** Each of these variables is an alternative collapsing of the corresponding \(X_{5j}\) into the dichotomy of “Succeed” or “Failed.” “Failed” occurs if the procedural analysis takes the value of “Strategic error,” “Tactical error,” or “Computational error.” “Success” signifies a correct response through an appropriate strategy.

\(X_{8j} \): **Response choice on Item j.** These variables are the actual values of subjects’ response choices: Mixture A juicier, Mixture B juicier, or equal. The parents of \(X_{8j}\) are \(X_{4j}\), strategy, and \(X_{5j}\), procedural analysis. That is, conditional on a particular choice of strategy and the way it is applied on a given item, what are the probabilities of each of the three potential response choices? Table 7 gives the conditional probability table for Item 17 as an example. Recall that whenever a strategy level is insufficient for an item’s structure, that strategy level for \(X_{4j}\) and “Success” for \(X_{5j}\) cannot co-occur. This fact is accounted for in the conditional probability matrix for \(X_{5j}\) given \(X_{4j}\), so the corresponding row in \(X_{8j}\)

\(^2\) One exception: two distinct strategies are classified as 1b; one is appropriate for Item 3 but the other is not.
is moot. Entries of equal probabilities appear as spaceholders. Other combinations that were not logically impossible but which few or no subjects exhibited were assigned conditional probabilities that reflected Béland’s judgement about likely outcomes, or, if there were no basis for such judgements, equal conditional probabilities.

[Insert Table 7 about here]

**X9j: “Objective” correctness of response on Item j.** These variables indicate whether the choices specified in X8j are in fact correct—regardless of how they have been reached. We refer to these as “objective” responses because they are typically the only observations that are available in standard multiple-choice “objective” educational tests. In that context they are thought of as “noisy” versions of the X6js. The conditional probabilities are logical: for “Correct,” the choice that happens to be correct for that item is assigned one and the other two are assigned zero; vice versa for “Incorrect.”

**Absorbing Evidence**

The construction of the network described in the preceding section exemplifies reasoning from causes to effects, as it were. The initial status shown as Figure 5 represents our state of knowledge about a new individual from the same population, beliefs about her likely responses to the sample items and the optimal stage x level we might expect to observe over the course of the twenty-item test. Once she begins to respond, we update our knowledge about observed variables directly, and about still unobserved variables probabilistically. This section offers some examples of how observations update beliefs, particularly with regard to X1, “optimal cognitive level,” and X2, “optimal stage x level.” We focus on some interesting contrasts among the strength and nature of various observations for inferring subjects’ cognitive levels.

Recall that these data provide two distinct pieces of evidence on each item, a response choice and an explanation. A first example illustrates a distinction between the value of evidence from the two. Figure 7 shows the network after an incorrect response has been observed to Item
The updated status of $X_{6,17}$, the "Structure understood?" variable for Item 17, indicates an 88% probability that this occurred because of an insufficient strategy and 12% due to inaccurate execution of a sufficient strategy, with probabilities of particular strategy levels shown in $X_{4,17}$, the "Item strategy" variable for Item 17. Initial beliefs for cognitive levels 1, 2, and 3 in $X_1$ of 8%, 45%, and 47% have shifted down to 13%, 54%, and 33% (c.f. Figure 5). Expectations for correct responses and understandings of Items 3 and 8 have also been downgraded. Figure 8 shows the additional updating that occurs if we learn this incorrect response was arrived at by a strategy at level 3b, the level isomorphic to the item. Probable explanation for the failure is 20% tactical error, 80% computational error. Belief about overall cognitive level is concentrated on Level 3, and expectations for correct responses to remaining items increase beyond their initial status.

As mentioned above, correct answers to multiple-choice items are typically taken as proxies for correct understandings in educational testing. Test developers avoid items with high "false positive" rates, or probabilities of correct answers by chance or by incorrect reasoning. Figure 9 reveals that Item 17 is just such an item. Of the subjects who responded with the correct choice, fewer than half did so with a strategy that accounted for the true structure of the item! In particular, a quarter of the correct responders employed a level 1b strategy: (3,5) is less juicy than (2,3) because (3,5) has more water. For this reason, a correct response on Item 17 shifts beliefs about optimal level upward only slightly. A correct explanation, on the other hand, would immediately establish certain belief at Level 3.

In contrast, Item 8 is a good multiple-choice item by test theoretic standards. Figure 10 shows that the overwhelming majority of subjects who answered correctly did so through a correct understanding of the equivalence-class structure of the item. Interestingly, posterior beliefs shift substantially to level 3 even though only a level 2
strategy is required for correct understanding. This is because nearly all the subjects whose optimal level was 3 understood the structure of Item 8, while less than half of those whose optimal level was 2 did. To further identify whether a correct responder had level 2 or level 3 as an optimal cognitive level would require additional information, such as checking the Item 8 explanation to see if it employed a level 3 strategy (if not, the probability for level 3 would be reduced but not eliminated), or presenting a level 3 item not so prone as Item 17 to false positives (an incorrect response would shift belief to level 2, a correct one to level 3). We note in passing that the second of these options is conditionally independent of the Item 8 choice, given optimal level, whereas the first is not. The DAG (Figure 4) indicates the potential confounding or overlap of information about optimal level from multiple aspects of a response to a given item, due to the presence of the shared "Item strategy" variables linking aspects of information from the same item. One avoids "double counting," or overinterpreting partially redundant information by acting as if it were independent, by properly accounting for the inferential structure of the observations, as demonstrated in this example.

[Insert Figures 9 and 10 about here]

The question of which observation to secure next is addressed by a series of "what if" experiments—a preposterior analysis, in Bayesian terminology. At a given state of knowledge, one can run through the values of a yet unobserved variable, summing the information (in terms of, say, reduced entropy or decreased loss) at each with weights proportional to their predicted probability under current beliefs. The next observation can then be selected to be optimal, in terms of, say, reducing expected loss or reducing expected entropy for a particular unknown variable. This is a straightforward application of statistical decision theory (Raiffa & Schlaifer, 1961).
Comments on the Example

This network provides a simple demonstration compared to the range of potential applications for probabilistic inference about cognitive student models. It does illustrate, however, probability-based reasoning built around structural relationships among cognitive strategies and the qualitatively different states of knowledge under a theory for the acquisition of proportional reasoning.

One of the limitations of this model is that it only provides an explanation of the individual's knowledge organization for a single ability. Consequently, one next step in development might be broadening the scope of the model to accommodate more than one ability—for example, proportional reasoning in a different domain, or something more disparate such as spatial visualization or short-term memory capacity. This can be accomplished by analyzing the structural relationships among individuals' state of learning in different domains. From the cognitive researcher's point of view, an interesting outcome of this study is that it opens up new avenues of exploration in the research of mechanisms and/or processes that lead to the construction of knowledge. Such efforts might create new perspectives for a test theory based on cognitive models. The inferential machinery explored here complements the skill lattice theory Haertel and Wiley (in press) propose as a basis for constructing educational achievement tests.

A more serious limitation is the treatment of subjects' cognitive state. Optimal level was operationalized in the network as the highest strategy level that a subject employed during the course of observation. This is appropriate for inferring the likelihood of a subject's highest level in the entire set knowing just a selected subset of responses. It only tells the whole story, however, under the assumption that a subject's likelihoods of response remained constant over the course of testing—that is, that a subject's toolkit of available cognitive strategies remains unchanged during testing. There is evidence that this is not the case. Cases have been observed in which a subject's previously highest level strategy proves inadequate for a subsequent item, the subject recognizes its inadequacy,
and, in response to the structure of the item, adapts or extends previous strategies or devises new concepts and strategies. Indeed, selecting an item most likely to provoke this kind of restructuring lies at the essence of cognitive-based instruction (Vosniadou & Brewer, 1987)!

The data from which the inference network described above was constructed would support an analysis of this phenomenon, and such work is currently in progress. Figure 11 sketches one direction in which the network described above might be extended to capture key aspects of it. Rather than a single variable expressing a subject’s cognitive status throughout the test, there is a distinct variable for each item presented. Cognitive status as it is in effect for Item \( j \) depends on the individual’s cognitive status as it was before the item was presented and on the structure of Item \( j \) itself. The probability that assimilation or accommodation may occur from this interplay is expressed in a new “cognitive processes” variable. We would expect probabilities of adaptive restructuring to be essentially zero when the structure of the item lies below the subject’s entering level and low when the item structure is far above her entering level, but maximal when the item lies just beyond what she has been able to handle up to that point.

[Insert Figure 11 about here]

**Discussion**

A host of practical issues must be addressed in exploring the applicability of probability-based inference, via inference networks, to cognitive assessment. We conclude by mentioning a number of them.

*More ambitious student models.* The proportional reasoning network discussed above has a very simple representation at its deepest level—a single “optimal level” variable entailing a class of available concepts and strategies. Our challenge was to model the structure of uncertain, partially redundant, sometimes conflicting evidence that observations
convey about the deep variable. A single deep variable is obviously too simple for many practical applications, and we must explore ways to implement student models with many descriptors of knowledge structures, multiple strategy options, and metacognitive and/or affective variables.

*The assumed completeness of the network.* The inference networks we have discussed are closed systems, which presume to account for all relevant possibilities; i.e., the space of student models is complete. In any application we can hope at best to model the key features distinguishing learners, certainly missing differences that will impact behavior. These differences are modelled as random variation. How does this affect inference? Can we build networks in such a way as to identify unexpected patterns, and to minimize resulting inferential errors?

*The nature of student models.* Our basic idea is to provide for probabilistic reasoning from observations to student models. This idea can be entertained for any type of student models, but certainly it will prove more useful for some types of student models than others. Characteristics of student models that need to be explored in this connection include model grain-size, and the distinctions between overlay vs. performance models (Ohlsson, 1986), and static vs. dynamic models.

- Grain-size concerns the level of detail at which to model students. As Greeno (1976) points out, “It may not be critical to distinguish between models differing in processing details if the details lack important implications for quality of student performance in instructional situations, or the ability of students to progress to further stages of knowledge and understanding.” The grain-size of our example was stage x level. A coarser model would address level only, while a finer model might further differentiate strategies within stages within levels.

- An “overlay” approach to diagnosing knowledge in the context of intelligent tutoring systems builds a representation of an expert’s knowledge base, and infers
from observed behavior where a student’s representation falls short (e.g., C. Frederiksen & Breuleux, 1989). A “performance model” attempts to specify correct and/or incorrect elements of knowledge and application rules in sufficient detail to solve the same problems the student is attempting (e.g., VanLehn, 1990). Our example was a probabilistic version of a simple performance model, as it provides predictions of response probabilities for all items for subjects at all modelled states.

- Static models assume a constant knowledge structure during the course of data-gathering; dynamic models expect, and attempt to model, changes in the learner along the way. The latter is obviously more ambitious, yet critical to applications such as ITSs in which learning is expected. White and J. Frederiksen’s (1987) QUEST system, for example, builds performance models in the domain of simple electrical circuits; the process of instruction is viewed as facilitating the evolution of models, successively shaping student understanding toward that of an expert. Kimball’s (1982) calculus tutor utilizes an approach that might be generalized: A student model is built under an assumption of statis during a problem, but the prior distribution for the next problem is modified to reflect the outcome of the experience and a reinforcement model. Our example was static; Figure 11 sketched one possible dynamic extension.

**Decision-making and prediction.** In the context of medical diagnosis, Szolovits and Pauker (1978, p. 128) point out the necessity of “...introducing some model of disease evolution in time, and dealing with treatment, as diagnosis is hard to divorce from therapy in any practical sense.” In the context of education, we are concerned with learning and instruction. The Bayesian inferential machinery, as a component of statistical prediction and decision theory, is natural for this task. What is required is to extend a network to prediction and decision nodes, and to incorporate utilities as well as probabilities into
decision rules. Andreassen, Jensen, and Olesen (1990) illustrate these ideas with a simple example from medical diagnosis. We must lay out the analogous extension in networks for cognitive assessment.

**Practical tools.** While the inference network approach holds promise for tackling class of problems in cognitive assessment, we are a long way from routinely engineering solutions to particular members of that class. This requires a methodological toolkit of generally applicable techniques and well-understood approaches. Building block models and heuristics are useful, for example, so that each application need not start from scratch. Foundational work on building-block models appears in Schum (1987). Work tailored to the kinds of observational settings and the kinds of psychological models anticipated in educational applications is required. And since simplifications of reality are inevitable, it is important to learn about the consequences of various model violations, and to develop diagnostic techniques for detecting serious ones.

### Conclusion

The modelling approach sketched in this paper was motivated by the following consideration:

Standard test theory evolved as the application of statistical theory with a simple model of ability that suited the decision-making environment of most mass educational systems. Broader educational options, based on insights into the nature of learning and supported by more powerful technologies, demand a broader range of models of capabilities—still simple compared to the realities of cognition, but capturing patterns that inform a broader range of instructional alternatives. A new test theory can be brought about by applying to well-chosen cognitive models the same general principles of statistical inference that led to standard test theory when applied to the simple model. (Mislevy, in press).
Probabilistic inference about cognitive student models via inference networks provides a potential framework for a more broadly based test theory. Exploiting conceptual and computational advances in statistical inference, the approach presents an opportunity to extend the achievements of model-based measurement to educational problems cast in terms of contemporary cognitive and educational psychology.
References


<p>| 1. | Mislevy and Verhelst’s (1990) mixture models for item responses when different examinees follow different solution strategies or use alternative mental models. |
| 2. | Falmagne’s (1989) and Haertel’s (1984) latent class models for Binary Skills. Students are modelled in terms of the presence or absence of elements of skill or knowledge, and observational situations demand various combinations of them. |
| 3. | Masters and Mislevy’s (in press) and Wilson’s (1989a) use of the Partial Credit rating scale model to characterize levels of understanding, as evidenced by the nature of a performance rather than its correctness. This incorporate into a probabilistic framework the cognitive perspective of Biggs and Collis’s (1982) SOLO taxonomy for describing salient qualities of performances. |
| 4. | Wilson’s (1989b) Saltus model for characterizing stages of conceptual development, which model parameterizes differential patterns of strength and weakness as learners progress through successive conceptualizations of a domain. |
| 5. | Yamamoto’s (1987) Hybrid model for dichotomous responses. This model characterizes an examinee as either belonging to one of a number of classes associated with states of understanding, or in a catch-all IRT class. The approach is useful when certain response patterns signal states of understanding for which particular educational experiences are known to be effective. |
| 6. | Embretson’s (1985) multicomponent models integrate item construction and inference within a unified cognitive model. The conditional probabilities of solution steps given a multifaceted student model are given by statistical structures developed in IRT. |
| 7. | Tatsuoka’s (1989) Rule space analyses uses a generalization of IRT methodology to define a metric for classifying examinees based on likely patterns of item response given patterns of knowledge and strategies. |</p>
<table>
<thead>
<tr>
<th>Medical Application</th>
<th>Educational Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observable symptoms, medical tests</td>
<td>Test items, verbal protocols, observers’ ratings, solution traces</td>
</tr>
<tr>
<td>Disease states, syndromes</td>
<td>States or levels of understanding of key concepts, available strategies</td>
</tr>
<tr>
<td>Architecture of interconnections based on medical theory</td>
<td>Architecture of interconnections based on cognitive and educational theory</td>
</tr>
<tr>
<td>Conditional probabilities given by physiological models, empirical data, expert opinion</td>
<td>Conditional probabilities given by psychological models, empirical data, expert opinion</td>
</tr>
</tbody>
</table>
**Table 3**
**Stages within Cognitive Levels**

<table>
<thead>
<tr>
<th>Level 1: Conceptual or preoperational</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Sole comparison of the number of juice glasses, the <strong>attribute</strong> in both pairs.</td>
</tr>
<tr>
<td>b. Appraisal of the dilution effect of the water on the final taste of juice. From this, the order of magnitude became a comparison of the number of water glasses, the <strong>complement</strong> in both pairs.</td>
</tr>
<tr>
<td>c. Construction of functional relations between the complementary terms in each pair, establishing <strong>between</strong> relations in the pair of <strong>within</strong> relations first constructed.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level 2: Concrete operational</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Use of the ratio “one glass of juice for one glass of water” to demonstrate that both terms within each pair were equal.</td>
</tr>
<tr>
<td>b. Joint multiplication of both terms within a pair or, otherwise, an operation of co-multiplication. (Scalar operator; e.g.,”Both drinks taste alike because there is one glass of juice for three glasses of water, which defines the ratio 1:3 in both pairs.”)</td>
</tr>
<tr>
<td>c. Relationships formed between both terms of each pair, when the first term was divided by the second. (Functional operator; e.g.,&quot;The ratio of two glasses of juice for six glasses of water is the same as one glass of juice for three glasses of water. Three times the ratio 1:3 equal three glasses of juice for nine glasses of water. Therefore both drinks taste alike.”)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level 3: Formal operational</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Either a scalar or functional operator in the <strong>between</strong> or the <strong>within</strong> relations.</td>
</tr>
<tr>
<td>b. Ratio treatment: The components of the relationships were the attribute and the complement. (E.g., “In Mixture A there are three glasses of juice for five glasses of water, a ratio of 9:15. In Mixture B the ratio is 10:15 juice to water. Therefore, Mixture B tastes juicier.”)</td>
</tr>
<tr>
<td>c. Fraction treatment: the components of the relationships were the attribute and the quantity of liquid. (E.g., “In Mixture A, of a total of 8 glasses, 3 contain juice, representing a fraction of 15/40. In Mixture B, of a total of 5 glasses, 2 were juice, representing a fraction of 16/40. Therefore, Mixture B tastes juicier.”)</td>
</tr>
</tbody>
</table>
Table 4
Conditional Probabilities of Stages within Cognitive Levels

<table>
<thead>
<tr>
<th>Level</th>
<th>a</th>
<th>b</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.000</td>
<td>.612</td>
<td>.388</td>
</tr>
<tr>
<td>2</td>
<td>.582</td>
<td>.345</td>
<td>.073</td>
</tr>
<tr>
<td>3</td>
<td>.145</td>
<td>.667</td>
<td>.188</td>
</tr>
</tbody>
</table>
Table 5
Conditional Probabilities of Strategies given Optimal Cognitive Levels

<table>
<thead>
<tr>
<th>Optimal Level</th>
<th>Strategy Level of Response</th>
<th>Ud.</th>
<th>1a</th>
<th>1b</th>
<th>1c</th>
<th>2a</th>
<th>2b</th>
<th>2c</th>
<th>3a</th>
<th>3b</th>
<th>3c</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Item 3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1a</td>
<td></td>
<td>.50</td>
<td>.50</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>1b</td>
<td></td>
<td>.08</td>
<td>.04</td>
<td>.88</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>1c</td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>.34</td>
<td>.64</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>2a</td>
<td></td>
<td>.01</td>
<td>.02</td>
<td>.37</td>
<td>.39</td>
<td>.21</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>2b</td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>.34</td>
<td>.54</td>
<td>.09</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>2c</td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>.39</td>
<td>.52</td>
<td>.06</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>3a</td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>.20</td>
<td>.74</td>
<td>.02</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>3b</td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>.02</td>
<td>.21</td>
<td>.02</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.71</td>
<td>.00</td>
</tr>
<tr>
<td>3c</td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>.18</td>
<td>.02</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.10</td>
<td>.65</td>
<td></td>
</tr>
<tr>
<td>(Item 8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1a</td>
<td></td>
<td>.50</td>
<td>.50</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>1b</td>
<td></td>
<td>.01</td>
<td>.04</td>
<td>.95</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>1c</td>
<td></td>
<td>.01</td>
<td>.02</td>
<td>.96</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>2a</td>
<td></td>
<td>.01</td>
<td>.02</td>
<td>.58</td>
<td>.04</td>
<td>.35</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>2b</td>
<td></td>
<td>.01</td>
<td>.02</td>
<td>.32</td>
<td>.01</td>
<td>.31</td>
<td>.33</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>2c</td>
<td></td>
<td>.01</td>
<td>.02</td>
<td>.06</td>
<td>.01</td>
<td>.24</td>
<td>.60</td>
<td>.06</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>3a</td>
<td></td>
<td>.01</td>
<td>.02</td>
<td>.11</td>
<td>.01</td>
<td>.08</td>
<td>.74</td>
<td>.02</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>3b</td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.41</td>
<td>.01</td>
<td>.01</td>
<td>.52</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>3c</td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.29</td>
<td>.01</td>
<td>.01</td>
<td>.07</td>
<td>.57</td>
<td></td>
</tr>
<tr>
<td>(Item 17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1a</td>
<td></td>
<td>.50</td>
<td>.50</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>1b</td>
<td></td>
<td>.07</td>
<td>.01</td>
<td>.92</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>1c</td>
<td></td>
<td>.04</td>
<td>.01</td>
<td>.94</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>2a</td>
<td></td>
<td>.03</td>
<td>.01</td>
<td>.43</td>
<td>.06</td>
<td>.47</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>2b</td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>.46</td>
<td>.01</td>
<td>.51</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>2c</td>
<td></td>
<td>.04</td>
<td>.01</td>
<td>.13</td>
<td>.01</td>
<td>.50</td>
<td>.00</td>
<td>.31</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>3a</td>
<td></td>
<td>.04</td>
<td>.01</td>
<td>.12</td>
<td>.03</td>
<td>.40</td>
<td>.00</td>
<td>.18</td>
<td>.22</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>3b</td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.04</td>
<td>.00</td>
<td>.01</td>
<td>.01</td>
<td>.90</td>
<td>.00</td>
</tr>
<tr>
<td>3c</td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
<td>.01</td>
<td>.01</td>
<td>.18</td>
<td>.75</td>
<td></td>
</tr>
</tbody>
</table>
Table 6
Conditional Probabilities of Procedural Analysis given Item Strategies

<table>
<thead>
<tr>
<th>Item Strategy</th>
<th>Success</th>
<th>Strategic Error</th>
<th>Tactical Error</th>
<th>Computational Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Item 3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ud</td>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>1a</td>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>1b</td>
<td>.75</td>
<td>.20</td>
<td>.05</td>
<td>.00</td>
</tr>
<tr>
<td>1c</td>
<td>.98</td>
<td>.00</td>
<td>.02</td>
<td>.00</td>
</tr>
<tr>
<td>2a</td>
<td>.85</td>
<td>.00</td>
<td>.15</td>
<td>.00</td>
</tr>
<tr>
<td>2b</td>
<td>.98</td>
<td>.00</td>
<td>.01</td>
<td>.00</td>
</tr>
<tr>
<td>2c</td>
<td>.97</td>
<td>.00</td>
<td>.02</td>
<td>.01</td>
</tr>
<tr>
<td>3a</td>
<td>.96</td>
<td>.00</td>
<td>.02</td>
<td>.02</td>
</tr>
<tr>
<td>3b</td>
<td>.98</td>
<td>.00</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>3c</td>
<td>.90</td>
<td>.00</td>
<td>.08</td>
<td>.02</td>
</tr>
<tr>
<td>(Item 8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ud</td>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>1a</td>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>1b</td>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>1c</td>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>2a</td>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>2b</td>
<td>.98</td>
<td>.00</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>2c</td>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>3a</td>
<td>.98</td>
<td>.00</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>3b</td>
<td>.98</td>
<td>.00</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>3c</td>
<td>.96</td>
<td>.00</td>
<td>.02</td>
<td>.02</td>
</tr>
<tr>
<td>(Item 17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ud</td>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>1a</td>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>1b</td>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>1c</td>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>2a</td>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>2b</td>
<td>.98</td>
<td>.00</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>2c</td>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>3a</td>
<td>.70</td>
<td>.00</td>
<td>.10</td>
<td>.20</td>
</tr>
<tr>
<td>3b</td>
<td>.95</td>
<td>.00</td>
<td>.01</td>
<td>.04</td>
</tr>
<tr>
<td>3c</td>
<td>.97</td>
<td>.00</td>
<td>.02</td>
<td>.01</td>
</tr>
</tbody>
</table>
Table 7
Conditional Probabilities of Item 17 Choice given Item Strategies and Procedural Analysis

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Procedural Analysis</th>
<th>Choice</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Undifferentiated</td>
<td>Success</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>Undifferentiated</td>
<td>Strategic Error</td>
<td>.13</td>
<td>.12</td>
<td>.75</td>
</tr>
<tr>
<td>Undifferentiated</td>
<td>Tactical Error</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>Undifferentiated</td>
<td>Computational Error</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>1a</td>
<td>Success</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>1a</td>
<td>Strategic Error</td>
<td>.98</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>1a</td>
<td>Tactical Error</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>1a</td>
<td>Computational Error</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>1b</td>
<td>Success</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>1b</td>
<td>Strategic Error</td>
<td>.23</td>
<td>.76</td>
<td>.01</td>
</tr>
<tr>
<td>1b</td>
<td>Tactical Error</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>1b</td>
<td>Computational Error</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>1c</td>
<td>Success</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>1c</td>
<td>Strategic Error</td>
<td>.01</td>
<td>.01</td>
<td>.98</td>
</tr>
<tr>
<td>1c</td>
<td>Tactical Error</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>1c</td>
<td>Computational Error</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>2a</td>
<td>Success</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>2a</td>
<td>Strategic Error</td>
<td>.03</td>
<td>.95</td>
<td>.02</td>
</tr>
<tr>
<td>2a</td>
<td>Tactical Error</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>2a</td>
<td>Computational Error</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>2b</td>
<td>Success</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>2b</td>
<td>Strategic Error</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>2b</td>
<td>Tactical Error</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>2b</td>
<td>Computational Error</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>2c</td>
<td>Success</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>2c</td>
<td>Strategic Error</td>
<td>.01</td>
<td>.01</td>
<td>.98</td>
</tr>
<tr>
<td>2c</td>
<td>Tactical Error</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>2c</td>
<td>Computational Error</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
</tbody>
</table>

(continued)
Table 7, continued
Conditional Probabilities of Item 17 Choice given Item Strategies and Procedural Analysis

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Procedural Analysis</th>
<th>Choice</th>
<th>Mixture A</th>
<th>Mixture B</th>
<th>Equal</th>
</tr>
</thead>
<tbody>
<tr>
<td>3a</td>
<td>Success</td>
<td></td>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
</tr>
<tr>
<td>3a</td>
<td>Strategic Error</td>
<td></td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>3a</td>
<td>Tactical Error</td>
<td></td>
<td>.80</td>
<td>.00</td>
<td>.20</td>
</tr>
<tr>
<td>3a</td>
<td>Computational Error</td>
<td></td>
<td>.50</td>
<td>.00</td>
<td>.50</td>
</tr>
<tr>
<td>3b</td>
<td>Success</td>
<td></td>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
</tr>
<tr>
<td>3b</td>
<td>Strategic Error</td>
<td></td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>3b</td>
<td>Tactical Error</td>
<td></td>
<td>.50</td>
<td>.00</td>
<td>.50</td>
</tr>
<tr>
<td>3b</td>
<td>Computational Error</td>
<td></td>
<td>.38</td>
<td>.00</td>
<td>.62</td>
</tr>
<tr>
<td>3c</td>
<td>Success</td>
<td></td>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
</tr>
<tr>
<td>3c</td>
<td>Strategic Error</td>
<td></td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>3c</td>
<td>Tactical Error</td>
<td></td>
<td>.90</td>
<td>.00</td>
<td>.10</td>
</tr>
<tr>
<td>3c</td>
<td>Computational Error</td>
<td></td>
<td>.70</td>
<td>.00</td>
<td>.30</td>
</tr>
</tbody>
</table>
Figure 1

Graphical Representations in the IRT Example
Which mixture will be more juicy—A, B, or both the same?

Figure 2

A Sample Juice-Mixing Task
Figure 3

Three Juice-Mixing Tasks
Figure 4
Graph of the Juice-Mixing Network
Figure 5

Initial Status, with Marginal Probabilities
Figure 6

Status Conditional on Optimal Level = 3b
Figure 7

Status Conditional on Item 17 Response Choice = Wrong
Figure 8

Status Conditional on Item 17 Response Choice = Wrong
and Item 17 Strategy = Level 3b
Figure 10

Status Conditional on Item 8 Response Choice = Right
Figure 11

Sketch for a Dynamic Inference Network
Distribution List

Dr. Terry Anderson
Educational Psychology
200C Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Beth Adelson
Department of Psychology
Rutgers University
Camden, NJ 08102

Dr. Robert Abras
21250 Research Parkway
Human Factors Division, Code 241
Naval Training Systems Center
Orlando, FL 32826

Technical Document Center
ALHGR-TDC
Wright-Patterson AFB
OH 45433-4503

Dr. Terry Allard
Code 1142C
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217-5000

Dr. Nancy Allen
Educational Testing Service
Princeton, NJ 08541

Dr. James A. Anderson
Department of Cognitive and Linguistic Sciences
Brown University
Box 1978
Providence, RI 02912

Dr. John R. Anderson
Department of Psychology
Carnegie-Mellon University
Schrenk Park
Pittsburgh, PA 15213

Dr. Nancy S. Anderson
Department of Psychology
University of Maryland
College Park, MD 20742

Dr. Stephen J. Andreoie, Chairman
College of Information Studies
Drexel University
Philadelphia, PA 19104

Dr. Gregory Angir
Educational Testing Service
Princeton, NJ 08541

Dr. Fipps Asahi
Graduate School of Management
Rutgers University
92 New Street
Newark, NJ 07102-1895

Edward Atkins
13760 Lakeview Ct.
Rockville, MD 20850

Dr. Michael E. Atwood
VYXDS
AI Laboratory
500 Westchester Avenue
White Plains, NY 10604

prof. dot. Bruno G. Bella
Uscita di ricerca di
tecnologie artificiale
Università di Milano
20123 Milano - via P. Sforza 23
ITALY

Dr. William M. Bart
Department of Psychology
University of Minnesota
Minneapolis, MN 55455

Dr. James L. Bejar
Law School Admissions
Service Box 49
Norcross, GA 30090-0040

Leo Bahrenshoef
United States Nuclear Regulatoy Commission
Washington DC 20555

Dr. William O. Berry
APOSBLN1, N. Blvd 410
Bolling AFB, DC 20332-6448

Dr. Monika Bernbaum
Educational Testing Services
Princeton, NJ 08541

Dr. Werner P. Berne
Personalamt der Bundeswehr
Köln-CLIENT 90
FEDERAL REPUBLIC OF GERMANY

Dr. John Block
Teachers College, Box 8
Columbia University
525 West 121th Street
New York, NY 10027

Dr. Michael Blackburn
Code 940
Naval Ocean Systems Center
San Diego, CA 92152-5000

Dr. Bruce Bixson
Defense Manpower Data Center
99 Pacific St.
Suite 155A
Monterey, CA 93943-3231

Dr. Kenneth R. Boff
AL/CMPH
Wright-Patterson AFB
OH 45433-4573

Dr. C. Alan Borrus
Department of Psychology
George Mason University
4400 University Drive
Fairfax, VA 22030

Dr. Gwyneth Boudoo
Educational Testing Service
Princeton, NJ 08541

Dr. J. C. Boudreau
Manufacturing Engineering Lab
National Institute of Standards and Technology
Gaithersburg, MD 20879

Dr. Gordon H. Bower
Department of Psychology
Stanford University
Stanford, CA 94305

Dr. Richard L. Broeren
HQ, USM/ECOM/MEPCT
2500 Green Bay Road
North Chicago, IL 60064

Dr. Robert Bruner
Code 232
Naval Training Systems Center
Orlando, FL 32826-3224

Dr. Robert Brunner
American College Testing Program
P. O. Box 148
Iowa City, IA 52243

Dr. Ann Brown
Graduate School of Education
University of California
EMST-4533 Tolman Hall
Berkeley, CA 94720

Dr. David V. Buxbaum
Department of Psychology
University of Hawaii
Mount Carmel, Hula, 91999

Dr. Gregory Cannell
CTP/Thomflan/McCowan-Hill
2500 Garden Road
Monterey, CA 93940

Dr. Gail Carpenter
Center for Adaptive Systems
111 Cummings St., Room 244
Boston University
Boston, MA 02215

Dr. Pat Carpenter
Carnegie-Mellon University
Department of Psychology
Pittsburgh, PA 15213

Dr. Edward Casselar
Educational Testing Services
Rosedale Road
Princeton, NJ 08541

Dr. Paul R. Chastellier
Perceptronics
1911 North Ft. Myer Dr.
Suite 800
Arlington, VA 22209

Dr. Micheleene Chi
Learning R & D Center
University of Pittsburgh
3009 O'Hara Street
Pittsburgh, PA 15260

Dr. Susan Chipman
Cognitive Science Program
Office of Naval Research
800 North Quincy St.
Arlington, VA 22217-5000

Dr. Raymond E. Christal
U.S. Laboratory for Atmospheric Research
AL/FRAMIL
Brooks AFB, TX 78235

Dr. William J. Closson
Institute for Research on Learning
2550 Hanover Street
Palo Alto, CA 94304

Dr. Norman C Offen
department of Psychology
U. of. Southern California
Los Angeles, CA 90089-1061

Dr. Paul Cobb
Purdue University
Eductional Building
W. Lafayette, IN 47907

Dr. Rodney Cooling
INM09, Basic Behavior and Cognitive Science Research
5600 Fisher Lane, Room A-013B
Parklawn Building
Rockville, MD 20897
CDR Frank C. Pehoe
Naval Postgraduate School
Code OR/PE
Monterey, CA 93943

Dr. Peter Pirofi
School of Education
University of California
Berkeley, CA 94720

Prof. Tommaso Poggio
Massachusetts Institute of Technology
Department of Electrical Engineering
Information Processing
Cambridge, MA 02139

Dr. Martha Poisson
Department of Psychology
University of Colorado Boulder, CO 80309-0344

Dr. Peter Poisson
University of Colorado Boulder, CO 80309-0344

Dr. Joseph Poole
ATTN: PER-HC
Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22335-5600

Gayle P. Poole - GDC-Telecomm and M
American Psychological Assoc.
1250 Uslee Street
Arlington, VA 22201

Mr. Paul S. Paul
Code U-23
Naval Surface Warfare Center
White Oak Laboratory
Silver Spring, MD 20903

Dr. Mark D. Peterson
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. James A. Persigo
Dept. of Computer Science
A. V. Williams Blvd.
University of Maryland College Park, MD 20742

Dr. J. Wesley Ragan
AFHRL/DI
Brooks AFB, TX 78235

Dr. Daniel Reisberg
Rensselaer College
Department of Psychology
Portland, OR 97202

Mr. Steve Reis
Department of Psychology
University of California Riverside, CA 92521

Dr. Brian Reiser
Department of Psychology
Green Hall
Princeton University
Princeton, NJ 08544

Dr. Lauren Remick
Learning R & D Center
University of Pittsburgh
3009 O'Hara Street
Pittsburgh, PA 15213

Dr. Gilbert Rice
Mail Stop E20-14
Grumman Aircraft Systems
Bethpage, NY 11714

Dr. Edmund Riesland
Dept. of Computer and Information Science
University of Massachusetts Amherst, MA 01003

Dr. Linda G. Roberts
Science, Education, and Transportation Program
Office of Technology Assessment
Congress of the United States
Washington, DC 20510

Dr. William B. Rose
Science Technology, Inc.
4703 Peachtree Corners Circle
Suite 200
Norcross, GA 30092

Dr. Louis Rosanno
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Donald Rubin
Statistics Department
Science Center, Room 408
1 Oxford Street
Harvard University
Cambridge, MA 02138

Dr. Fumio Sato
Department of Psychology
University of Tennessee
310B Austin Peay Bldg.
Knoxville, TN 37996-0900

Dr. Mark Schaefer
ENI International
333 Ravenswood Ave.
Room BS-131
Mento Park, CA 94025

Dr. Walter Schneider
Learning R&D Center
University of Pittsburgh
3039 O'Hara Street
Pittsburgh, PA 15260

Dr. Alan H. Schoenfeld
University of California
Department of Education
Berkeley, CA 94720

Dr. Mary Schuetz
4100 Parkside
Carbondale, CA 92008

Dr. Myra F. Schwartz
Director
Neuropsychology Research Lab
Moss Rehabilitation Hospital
1200 West Tabor Road
Philadelphia, PA 19141

Dr. Robert J. Szalai
US Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333

Dr. Colleen M. Selfert
Department of Psychology
University of Michigan
339 Packard Road
Ann Arbor, MI 48104

Dr. Terrence J. Sejnowski
Professor
The Salk Institute
P. O. Box 8500
San Diego, CA 92138-9216

Mr. Robert Simmons
NIST Ebon Hall
Department of Psychology
University of Minnesota
Minneapolis, MN 55455-0244

Dr. Valerie L. Shiler
Department of Industrial Engineering
State University of New York
342 Lawrence D. Bell Hall
Buffalo, NY 14260

Mr. Richard J. Stelwien
Graduate School of Education
University of California
Santa Barbara, CA 93106

Mr. Kathleen Steinmann
Educational Testing Service
Princeton, NJ 08541

Dr. Colin Sheppard
Command and Control Dept.
Defense Research Agency
Hartsfield Air Base
Portsmouth Haas P404A1
UNITED KINGDOM

Dr. Kazuo Shigesu
T-4-24 Kugenuma-Kitan
Pujwatu 251
JAPAN

Dr. Randall Shumaker
Naval Research Laboratory
Code 5520
455 Overlook Avenue, SW
Washington, DC 20375-5000

Scientific Director
Naval Health Research Center
P.O. Box 8122
San Diego, CA 92138-9174

Dr. Edward Silver
DRDC
University of Pittsburgh
3039 O'Hara Street
Pittsburgh, PA 15260

Dr. Zita M. Simatia
Director, Manpower & Personnel Research Laboratory
US Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Dr. James E. Singer
Department of Medical Psychology
Uniformed Services Univ. of the Health Sciences
4001 Jones Bridge Road
Bethesda, MD 20814-7799

Dr. Don E. Stroh
Computing Sciences Department
The University
Aberdeen A119 3PK
Scotland

UNITED KINGDOM

Dr. Robert Stallz
Naval Ocean Systems Center
Code 443
San Diego, CA 92152-5000

Dr. Richard E. Snow
School of Education
Stanford University
Stanford, CA 94305
<table>
<thead>
<tr>
<th>Name</th>
<th>Address</th>
<th>Phone</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr. Judy Sprey</td>
<td>ACT P.O. Box 148 Iowa City, IA 52243</td>
<td>319-335-4312</td>
<td><a href="mailto:jdsprey@act.com">jdsprey@act.com</a></td>
</tr>
<tr>
<td>Dr. Bruce D. Steinberg</td>
<td>Cary College McKinnon, IA 42126</td>
<td>315-794-9212</td>
<td><a href="mailto:bsteinberg@ccac.edu">bsteinberg@ccac.edu</a></td>
</tr>
<tr>
<td>Dr. Martha Stokking</td>
<td>Educational Testing Service Prinston, NJ 08541</td>
<td>609-981-6226</td>
<td><a href="mailto:mstokking@ets.org">mstokking@ets.org</a></td>
</tr>
<tr>
<td>Dr. William Stout</td>
<td>Department of Statistics University of Illinois 101 Main Hall 725 South Wright St. Champaign, IL 61800</td>
<td>217-444-2194</td>
<td><a href="mailto:wstout@illinois.edu">wstout@illinois.edu</a></td>
</tr>
<tr>
<td>Dr. Eiko Tanaka</td>
<td>Educational Testing Service Mall St Dy 03-7 Prinston, NJ 08541</td>
<td>609-258-3700</td>
<td><a href="mailto:etanaka@ets.org">etanaka@ets.org</a></td>
</tr>
<tr>
<td>Dr. David Thissen</td>
<td>Psychometric Laboratory CB 3270, Davie Hall University of North Carolina Chapel Hill, NC 27599-3270</td>
<td>919-515-2520</td>
<td><a href="mailto:dthissen@unc.edu">dthissen@unc.edu</a></td>
</tr>
<tr>
<td>Mr. Thomas J. Thomas</td>
<td>Federal Express Corporation Human Resource Development 3013 Director Row, Suite 561 Memphis, TN 38131</td>
<td>901-358-5555</td>
<td><a href="mailto:jthomas@fex.com">jthomas@fex.com</a></td>
</tr>
<tr>
<td>Mr. Gary Thomason</td>
<td>University of Illinois Educational Psychology Champaign, IL 61801</td>
<td>217-333-6688</td>
<td><a href="mailto:gthomason@illinois.edu">gthomason@illinois.edu</a></td>
</tr>
<tr>
<td>Chair, Department of</td>
<td>Psychology University of Maryland, Baltimore County Baltimore, MD 21228</td>
<td>410-704-3960</td>
<td></td>
</tr>
<tr>
<td>Dr. David VanLehn</td>
<td>Learning Research &amp; Development Ctr. University of Pittsburgh 3909 O'Hara Street Pittsburgh, PA 15260</td>
<td>412-648-7600</td>
<td>dv <a href="mailto:analeh@pitt.edu">analeh@pitt.edu</a></td>
</tr>
<tr>
<td>Dr. Fred L. Vaziri</td>
<td>New Personnel R&amp;D Center San Diego, CA 92152-6800</td>
<td>619-535-6100</td>
<td><a href="mailto:lvaziri@eds.com">lvaziri@eds.com</a></td>
</tr>
<tr>
<td>Dr. Jerry Vogt</td>
<td>Department of Psychology St. Norbert College De Pauw, WI 54115-2999</td>
<td>715-744-4111</td>
<td><a href="mailto:jvogt@stnorbert.edu">jvogt@stnorbert.edu</a></td>
</tr>
<tr>
<td>Dr. Jacques Vomachen</td>
<td>Department of Psychology Geneva Switzerland</td>
<td>412-624-7900</td>
<td><a href="mailto:jvomachen@ets.org">jvomachen@ets.org</a></td>
</tr>
<tr>
<td>Dr. Howard Weiner</td>
<td>Educational Testing Service Prinston, NJ 08541</td>
<td>609-981-6226</td>
<td><a href="mailto:hweiner@ets.org">hweiner@ets.org</a></td>
</tr>
<tr>
<td>Elizabeth Wald</td>
<td>Office of Naval Technology Code 227 800 North Quincy Street Arlington, VA 22217-5008</td>
<td>703-696-2600</td>
<td><a href="mailto:elizabethwald@onl.navy.mil">elizabethwald@onl.navy.mil</a></td>
</tr>
<tr>
<td>Dr. Michael T. Weiner</td>
<td>University of Wisconsin-Milwaukee Educational Psychology Dept. Box 413 Milwaukee, WI 53201</td>
<td>414-229-4652</td>
<td><a href="mailto:mweiner@wisc.edu">mweiner@wisc.edu</a></td>
</tr>
<tr>
<td>Dr. Ming-Mei Wang</td>
<td>Educational Testing Service Mall St Dy 03-7 Prinston, NJ 08541</td>
<td>609-258-3700</td>
<td><a href="mailto:mwang@ets.org">mwang@ets.org</a></td>
</tr>
<tr>
<td>Dr. Thomas A. Warm</td>
<td>FFA Academy P.O. Box 25065 Oklahoma City, OK 73125</td>
<td>405-521-4060</td>
<td><a href="mailto:twarm@ffacademy.org">twarm@ffacademy.org</a></td>
</tr>
<tr>
<td>Dr. David J. Weiss</td>
<td>University of Minnesota 75 E. River Road Minneapolis, MN 55455-8544</td>
<td>612-624-5544</td>
<td><a href="mailto:dweiss@umn.edu">dweiss@umn.edu</a></td>
</tr>
<tr>
<td>Dr. Douglas Wessel</td>
<td>Code 15 Navy Personnel R&amp;D Center San Diego, CA 92152-6800</td>
<td>619-553-1851</td>
<td><a href="mailto:dwessel@nRewdcenter.com">dwessel@nRewdcenter.com</a></td>
</tr>
<tr>
<td>Mr. John J. Young</td>
<td>National Science Foundation Room 320 800 G Street N.W. Washington, DC 20550</td>
<td>202-326-2060</td>
<td><a href="mailto:jjyoung@nsf.gov">jjyoung@nsf.gov</a></td>
</tr>
<tr>
<td>Dr. Michael Yan</td>
<td>Dept. of Computer Science University of Emmer Prince of Wales Road Emmer EX44PT England</td>
<td>011-886-8888</td>
<td><a href="mailto:muyan@computer.com">muyan@computer.com</a></td>
</tr>
<tr>
<td>Dr. Frank A. Yersi</td>
<td>Dept. of Education Catholic University Washington, DC 20084</td>
<td>202-624-2400</td>
<td><a href="mailto:fyersi@catholic.edu">fyersi@catholic.edu</a></td>
</tr>
<tr>
<td>Dr. Wendy Young</td>
<td>CTC/McGraw Hill Del Monte Research Park Monterey, CA 93940</td>
<td>831-373-2222</td>
<td><a href="mailto:wyoung@mcgrawhill.com">wyoung@mcgrawhill.com</a></td>
</tr>
<tr>
<td>Dr. Joseph L. Young</td>
<td>National Science Foundation Room 320 800 G Street N.W. Washington, DC 20550</td>
<td>202-326-2060</td>
<td><a href="mailto:jjyoung@nsf.gov">jjyoung@nsf.gov</a></td>
</tr>
<tr>
<td>Dr. Jerry Vost</td>
<td>Department of Political Science University of Connecticut 905 North Main Ave. Stamford, CT 06901</td>
<td>203-777-5000</td>
<td><a href="mailto:jvost@uconn.edu">jvost@uconn.edu</a></td>
</tr>
<tr>
<td>Dr. Mark Wilson</td>
<td>School of Education University of California Berkeley, CA 94720</td>
<td>510-642-0440</td>
<td><a href="mailto:mwilson@uc.edu">mwilson@uc.edu</a></td>
</tr>
<tr>
<td>Dr. Eugene Wray</td>
<td>Department of Psychology Emory University Atlanta, GA 30322</td>
<td>404-727-1331</td>
<td><a href="mailto:ewray@emory.edu">ewray@emory.edu</a></td>
</tr>
<tr>
<td>Dr. Robert A. Wiener</td>
<td>U.S. Army Institute for the Behavioral and Social Sciences 3001 Eisenhower Avenue Alexandria, VA 22333-5000</td>
<td>703-893-1923</td>
<td><a href="mailto:rwiener@iibs.gov">rwiener@iibs.gov</a></td>
</tr>
<tr>
<td>Dr. Martin F. Walch</td>
<td>PIERER/BIC 99 Pacific St. Suite 4536 Monterey, CA 93940</td>
<td>831-373-2222</td>
<td><a href="mailto:mwalch@mcgrawhill.com">mwalch@mcgrawhill.com</a></td>
</tr>
<tr>
<td>Dr. Martin C. Witteck</td>
<td>Graduate School of Education Univ. of Calif. Los Angeles Los Angeles, CA 90024</td>
<td>310-825-0600</td>
<td><a href="mailto:mwitteck@ucla.edu">mwitteck@ucla.edu</a></td>
</tr>
<tr>
<td>Mr. John H. Wolfe</td>
<td>Navy Personnel R&amp;D Center San Diego, CA 92152-6800</td>
<td>619-553-1851</td>
<td><a href="mailto:jwolfe@navy-rdcenter.com">jwolfe@navy-rdcenter.com</a></td>
</tr>
<tr>
<td>Dr. Ernesto Yamamoto</td>
<td>Educational Testing Service Rosenele Road Prinston, NJ 08541</td>
<td>609-258-3700</td>
<td><a href="mailto:eyamamoto@ets.org">eyamamoto@ets.org</a></td>
</tr>
<tr>
<td>Dr. Dennis Yan</td>
<td>University of Wisconsin-Milwaukee Educational Psychology Dept. Box 413 Milwaukee, WI 53201</td>
<td>414-229-4652</td>
<td><a href="mailto:dyan@wisc.edu">dyan@wisc.edu</a></td>
</tr>
</tbody>
</table>