Running head: Knowledge Space Modelling

Knowledge Space Modelling of Inductive Reasoning in an Intelligence Test

Jochen Musch

Rheinische Friedrich-Wilhelms-Universität Bonn

Dietrich Albert

Karl-Franzens-Universität Graz

Yellow Paper – Published Online December 2021

Abstract

The question of defining „intelligence“ often has been answered by referring to the competencies and skills which are necessary to perform intelligence tests. A prominent instrument for measuring general intelligence is the Advanced Progressive Matrices (APM) test of Raven. However, which skills are necessary to pass this test? Previous studies (e.g., Carpenter, Just and Shell, 1990; Vodegel Matzen, 1994) have identified important components that influence item difficulty: the number of necessary operations, their difficulty ranking and the difficulty of the material. On the basis of Knowledge Space Theory (Falmagne, Koppen, Villano, Doignon and Johannesen, 1990) in the formulation of Albert and Held (1994), item characteristics and inter-individual differences in performance were modelled in terms of these components. The order relations that best described the group performance data of a large ($N = 1015$) sample of Kratzmeier (1976) were identified and tested against two sets of individual answer patterns ($N = 44$ and $N = 41$). The results are discussed with respect to the theory and assessment of intelligence.

Fluid or analytic intelligence is, according to Cattell (1963, 1971), an expression of the level of complexity of relationships which an individual can perceive and act upon when he does not have recourse to information already stored in memory. One of the most prominent measures of fluid intelligence are the Advanced Progressive Matrices (APM; Raven, 1962; Raven, Court & Raven, 1983). Being the most complex version of the Raven Progressive Matrices, the APM are primarily intended for use with gifted subjects. A main objective of their development was to avoid ceiling effects that were often obtained when administering the Standard Progressive Matrices to high school and university populations (Raven, 1962).

As evidenced by a nonmetric scaling study of Snow, Kyllonen and Marshalek (1984), the APM hold a central position among psychometric tests of analytic intelligence. The high correlations between individual differences in the Raven test with those in other complex cognitive tests (Jensen, 1987) supply further evidence for the centrality of the test.

Maybe due to these properties, the APM are widely used in research, clinical and applied settings (e.g., Arthur, Barrett & Doverspike, 1990; Neubauer, Freudenthaler & Pfurtscheller, 1995; Pajarez & Kranzler, 1995, etc.). Due to its nonverbal format, the test is also a popular research tool whenever it is required to avoid educational bias and to minimize the influence of language on intelligence testing (Court, 1991). A cognitive analysis of the factors underlying performance on this test should have significant implications and might be of great practical importance in all domains in which the test is used. Moreover, because of its centrality one may expect that a theory of the solution process in the APM would also help to gain insight into many similar tests of analytic intelligence.

- Figure 1 about here -

All items of the APM are geometrical analogy problems of this format: nine entries are arranged in a 3 x 3 matrix. All entries consist of a composition of elements such as geometrical objects and patterns which differ in several attributes, e.g. number, spatial orientation, and shading. In each item, the bottom right-hand entry is missing, and its appearance is to be inferred from the relationships that exist rowwise and columnwise between the remaining entries. For each problem, there is a set of eight response alternatives. One of the alternatives depicted below of the problem represents the missing entry; the others are distractors. It is important to note that in the case of the APM, contrary to other domains such as chess problems (Albert, Schrepp & Held, 1994), the assumption seems justified that the sequence in which the rules that determine the missing entry are dealt with is irrelevant.

Unfortunately and somewhat surprisingly, despite a long history of research on the APM, the issue of what it means to perform well on the test is still not satisfactorily resolved. The answer to the question about the nature of intelligence has too often been Boring's (1923) dictum „Intelligence is what an intelligence test measures". We wanted to know: What exactly are the factors that in combination influence problem complexity in the APM, and how can they be modelled? What are the individual differences that make a difference between a high and a low score on this test of inductive reasoning?

An important step towards a better understanding of the APM was presented previously in this journal by Carpenter, Just and Shell (1990). These authors conducted a process analysis of the Raven test and suggested a theoretical account of what it means to perform well on a classic test of inductive reasoning like the APM. Based on item, protocol, and eye-movement analyses as well as on correlational and simulation studies, they formulated a theory concerned with what the APM actually measure.

According to Carpenter et al. (1990), the Raven test measures goal management ability, i.e. the ability to decompose problems into manageable sub-problems and to reassemble the various solutions into a solution of the problem as a whole. Their post-hoc analysis indicated three factors that contribute to the complexity of an item: (i) the number of sub-problems, (ii) the nature of the sub-problems (the kind of the solution rule), and (iii) the distinctiveness of the sub-problems. These same ingredients were confirmed as determinants of item difficulty by Vodegel Matzen (1994). In both investigations, lower scoring participants tended to fail to solve all sub-problems necessary for a solution due to a lower goal management ability. In particular, lower scoring participants were likely to omit multiple rules and failed to solve the more complex problems.

The results of Carpenter et al. (1990) and Vodegel Matzen (1994) can be formulated as follows: a participant masters a specific item in the test if and only if he or she possesses all the skills that are required for the solution of the respective item. For every single item in the test, a list of skills (competencies, cognitive abilities, operations, pieces of knowledge, etc.) assumed to be related to the solution of the problem can be specified (cf. Sternberg, 1977; Pellegrino & Glaser, 1979; Whitely, 1980).

Another approach not mentioned in Carpenter et al. (1990) that was also used for modelling performance on intelligence test items is the linear logistic version of the Rasch model (cf. Fischer, 1983; Formann, 1983; Formann & Piswanger, 1979; Hornke & Habon, 1984a,b; Keldermann & Rijkes, 1994; Nährer, 1980). This approach is capable of modelling both individual differences and task components. However, results were not very satisfying. Out of 42 items, Formann (1973; Fischer & Formann, 1982) deleted 9 items because they did not fit the model. Nevertheless, the attempt to explain difficulty as a function of item structure was not successful; all Likelihood ratio tests turned out to be significant. Nährer (1980) and Hornke and Habon (1984b) did not succeed in predicting item difficulty as a linear combination of the identified components either (cf. Formann, 1982). Moreover, it was never tried to model the much more complex APM items rather than the relatively simple Standard Progressive Matrices (SPM; e.g., Kelderman & Rijkes, 1994).

Another attempt at modelling the APM items also not mentioned by Carpenter et al. (1990) is the factor analytic approach. The reason for this might be that despite continuing efforts to determine the dimensionality of the APM, the results until now are rather equivocal and inconclusive. Various methodologies led to the extraction of 1 to 14 factors depending on choice of extraction method and correction technique. However, most of the factors found could not be replicated (Alderton & Larson, 1990; Arthur & Woehr, 1993; Dillon, Pohlmann & Lohmann, 1981). Results depended heavily on subject population and the selection of methodology (cf. McNemar, 1951; Guilford, 1952; Sternberg, 1977). However, it seems quite questionable on the basis of the factor analytic results whether performance on the APM can be reduced to one or a few determinants. For instance, a one-factor solution like that of Alderton & Larson (1990), accounting for only 34% of total variance, or a two-factor solution like that of Dillon et al. (1981), accounting for 43% of total variance, casts serious doubt on whether indeed all factors contributing to item difficulty were identified. Sternberg (1977) concluded that *„factor analysis is not an appropriate method for discovering the components underlying intelligence“* (p.29).

What is still lacking in the investigation of the APM is a formal description of the task characteristics that influence item difficulty and a comprehensive theory of how individuals differ in terms of the processes that are necessary to solve an item. We felt that the *Theory of Knowledge Spaces* which was presented previously in this journal by Falmagne, Koppen, Villano, Doignon and Johannesen (1990; cf. Doignon & Falmagne, 1985) is a promising new approach to this problem. Synthesizing the two approaches, we therefore modelled the demand components of the APM suggested by Carpenter et al. (1990) according to the the Theory of Knowledge Spaces (Falmagne et al., 1990). This synthesis allows to model individual differences in terms of the ability to deal with the problem components specific to each item. Our assumption was that success on the solution process is determined by the question of whether or not a test participant can master all specific components of a test item. We felt that a formal description of the participant’s competence states which can be observed empirically has several advantages compared to previous approaches.

First, the Theory of Knowledge Spaces formulates explicitly the link between demand components or (latent) skills and the observable solution pattern (see Albert & Lukas, 1999, for a recent overview). This connection to observable solutions of problems leads to falsifiable predictions about non-admissible solution patterns and thereby provides a tool for evaluating a cognitive theory about analytic intelligence based on a thorough item analysis. The componential approach specifies the prerequisites which are necessary for the solution of single items: it „gets inside“ the solution process. The principle is: try to formulate your theory about analytic ability explicitly enough to predict which solution patterns are compatible with it and which are not (Lukas & Albert, 1993). One goal of our study therefore was to answer the question of how the process of item solution can be meaningfully broken down into the participant’s ability to deal with the set of cognitive operations necessary for the solution of each item.

Second, a formal model of participant’s performance is suited to provide much more detailed diagnostic information about the test participants than the mere number of items solved by each subject. It also offers the chance to employ efficient, adaptive diagnostic procedures.

Third, whereas the number of factors that can be extracted in factor analysis depends heavily on sample homogeneity, competence structures can be formulated independent of sample properties. Moreover, rather than merely scaling observed differences as factor analysis does, the Theory of Knowledge Spaces allows the rigorous testing of substantive theories on the basis of individual answer patterns.

We will start with a brief review of the results of previous investigations into the APM that led to the identification of the components on which the APM items are based. Prior to formulating a model of performance in the APM, it will also be necessary to provide a brief introduction to the Theory of Knowledge Spaces.

The components of the APM

Several studies of Carpenter et al. (1990), Vodegel Matzen (1994) and Vodegel Matzen, van der Molen and Dudink (1994) indicate that three factors contribute to the complexity of the problems of the Raven test: (i) the number of rules, (ii) the kind of the solution rule, and (iii) the distinctiveness of the elements of the rule operations.

Number of rules

One major result of the study of Carpenter et al. (1990) was the finding that the problems in the APM vary in the number of rules that have to be detected in order to arrive at a correct solution. In order to detect these rules, it is necessary to generate subgoals in working memory, record the attainment of subgoals, and set new subgoals as others are attained (Carpenter et al., 1990). This procedure was shown to impose a load on working memory that increases with increasing number of rules (Carpenter et al., 1990; Vodegel Matzen, 1994). The importance of goal management in working memory is also shown by the fact that the correlation between performance on the Tower of Hanoi problem (which is known to involve extensive goal management; Egan & Greeno, 1974) and performance in the APM is close to the test-retest reliability of .77 typically found for the Raven test (Carpenter et al., 1990). Converging evidence for the involvement of goal management abilities was provided by Vodegel Matzen (1994).

The induction of each rule was found to consist of many small steps, reflected in the pairwise comparison of elements in adjoining entries (Carpenter et al., 1990). The incremental nature of the solution process was also apparent in eye-movement analyses which revealed multiple repeated scans of rows and columns and repeated fixations of pairs of related entries (Carpenter et al, 1990; Rhenius & Heydemann, 1984). In problems containing more than one rule, the rules are usually described one at a time in verbal reports, with long intervals between rule descriptions, suggesting that they are induced one at a time.

Because 45% of the variance in error rate was accounted for by the number of rules involved, a finding that was replicated by Vodel Matzen (1994), Carpenter and his coworkers concluded that the number of rules involved is the most important determinant of item difficulty. The first component we decided to include in our model therefore is the number of rules (N) involved in an item. Since the number of rules varies from one to four, the component N has four possible attributes: N = {*1,2,3,4*}. We assume that a linear order is defined on these attributes: a higher number of rules is assumed to make a problem more difficult.

Rule Taxonomy

Carpenter et al. (1990) found that five different operations accounted for most of the items in the subset of APM items they tried to classify in their taxonomy. In an analysis of the relations used in figural analogies (both 2 x 2 and 3 x 3 matrices) from 166 different intelligence tests, Jacobs and Vandeventer (1972) also found that relatively few, namely, twelve relations accounted for most of the problems. Similar taxonomies for various tests of analytic intelligence were provided by Ward and Fitzpatrick (1973), Formann (1973), and Hornke and Habon (1984a). Within all prior taxonomies, however, several problems of the test were unclassifiable and were therefore excluded from the analyses. It was our aim to provide a taxonomy that is able to account for *all* of the test items. Altogether, there are 46 items in the APM: 10 items from Set I (Item 3-12; the first two items are reserved for instruction purposes) and 36 items from Set II (Item 1-36). In order to model *all* of these items, we used a synopsis of the rules identified by Carpenter et al. (1990), Jacobs and Vandeventer (1972), Vodegel Matzen (1994) and Ward and Fitzpatrick (1973). We found that it was possible to describe all 46 items using seven elementary operations most of which are common to all of the above mentioned taxonomies. The seven rules that govern the variation among the items of the APM are:

1. Constant in a row (CR; Carpenter et al., 1990). The same value of an attribute occurs throughout a row (cf. Jacobs & Vandeventer, 1972; Vodegel Matzen, 1994; Ward & Fitzpatrick, 1973).

Example: a a a
b b b
c c ?

2. Quantitative pairwise progression (PP; Carpenter et al., 1990). A quantitative increment or decrement occurs between adjacent entries in an attribute such as size, position, or number (cf. Jacobs & Vandeventer, 1972; Vodegel Matzen, 1994; Ward & Fitzpatrick, 1973).

Example: a b c
d e f
g h ?

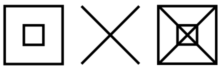
3. Distribution of two values (D2; Carpenter et al., 1990). Two values of an attribute are distributed through a row; the third value differs (cf. Vodegel Matzen, 1994).

Example: a a b
a b a
b a ?

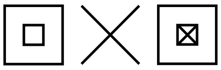
4. Distribution of three values (D3; Carpenter et al., 1990). Three values of an attribute are distributed in a permutative way through a row (cf. „elements of a set“, Jacobs & Vandeventer, 1972; Vodegel Matzen, 1994; „latin square“, Ward & Fitzpatrick, 1973).

Example: a b c
b c a
c a ?

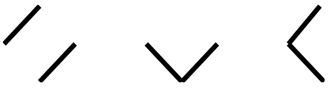
5. Figure addition (FA; Carpenter et al., 1990). A figure or element from one colum is added to another figure to produce the third (cf. Jacobs & Vandeventer, 1972; Vodegel Matzen, 1994).

Example: 

6. Figure subtraction (FS). A figure or element from one column is superimposed to another element to produce the third (cf. „figure subtraction“, Carpenter et al., 1990; Vodegel Matzen, 1994).

Example: 

7. Exclusive-OR (XO). This rule is equivalent to the „unique addition“ rule of Jacobs & Vandeventer (1972). It is expected to clearly exceed all other rules with respect to its difficulty: two attributes in a row are combined in the exclusive-or fashion of Boolean algebra. The following example is taken from item 36 of Set II (the most difficult of all items; parts of the item that belong to another rule are omitted to enhance clarity of presentation):

Example: 

The lines of a diamond are governed by the following rule: if there is a line in the respective segment of *exactly one* of the first two entries in a row, there is also a line in the respective segment of the third entry in the row. If, however, there are lines in *both* segments of the first two entries in a row *or* there are *no* lines in either segments of the first two entries, there is no line in the respective segment in the third entry of the row. The following truth table (1 = attribute present, 0 = attribute not present) applies:

$$1\ 1 \Rightarrow 0$$

$$1\ 0 \Rightarrow 1$$

$$0\ 1 \Rightarrow 1$$

$$0\ 0 \Rightarrow 0$$

Simulation studies (McClelland & Rumelhart, 1988) and empirical investigations (Lachnit, 1992; Kinder & Lachnit, 1994) provide evidence that the exclusive-or conjunction („unique addition“ in the terminology of Jacobs & Vandeventer, 1972) surpasses other rules in its difficulty. Converging evidence for the exceptional status of the exclusive-or conjunction with regard to its difficulty provide the classical works in concept learning and problem solving of Neisser and Weene (1962) and Haygood and Bourne (1965; cf. Neubauer, 1990a). The second component that we decided to include in our model therefore is the difficulty of the most difficult type of rule (T) involved in an item. We distinguished between items that involve the application of an exclusive-or rule (X) and items that consist of rules of (o)ther types only (O). Thus, the component T has two possible attributes: T = $\{X,O\}$, and we define the following difficulty order: $X > O$.

Material attributes

The items of the APM differ in the distinctiveness of the underlying sub-problems: the rules operate on material attributes of different detectability. Because of the problem of conceptually segmenting different rules, „the correspondence-finding process is a subtle source of difficulty“ (Carpenter et al., 1990, p. 410). Protocol analysis revealed that one of the main heuristics for correspondence finding is the „matching-names“ heuristic („matching-category heuristic“ in the terminology of Vodegel Matzen, 1994). This heuristic is based on the hypothesis that elements belonging to a category of characteristics or attributes with the same name (e.g. colour) should be grouped together and are governed by the same solution rule. Classification or categorization of objects is easiest if objects belong to known categories, so that they possess easily accessible characteristics which can be compared with one another. An alternative explanation for the same fact is that complex visual stimuli are harder to encode and keep in mind than simple ones (Vodegel Matzen, 1994). Because visual information is harder to maintain in short term memory, visual material is preferably stored acoustically, through labelling it with a name (Posner & Mitchell, 1967). Simple stimuli can be tagged with one name (e.g., „triangle"), whereas more complex visual stimuli need to be described in a sentence. Acoustically encoding and maintaining complex visual stimuli is therefore more difficult than the encoding of more simple stimuli. The extra demand that complex material puts on the working memory might prove fatal for the performance of a problem solver who is already allocating all resources to the solution process itself (Vodegel Matzen, 1994). In line with this reasoning, Vodegel Matzen (1994) found that providing more cues for the categorization of elements (cf. Rosch, 1977) and thereby facilitating correspondence finding results in lower error rates.

Some of the material attributes in the APM items are relatively easy to detect. They consist of geometrical figures (such as triangles, squares, circles, crosses etc) and patterns (lines, shadings, ornaments etc). Another material attribute upon which a rule can operate is the number of objects itself. For example, the number of circles in each entry can increase (pairwise progression). What is common to all these material attributes (geometrical figures, patterns, and the number of objects) is their relatively high salience.

However, there is another attribute that is rather difficult to detect, namely, spatial order. For example, the angle between the horizontal axis and the orientation of the shading of objects or the spatial orientation of a triangle can be the attribute upon which a rule operates. Such attributes are less salient, and because of their complex visual nature they presumably have to be stored acoustically (Posner & Mitchell, 1967).

The third component we include in our model therefore is the difficulty of the (m)aterial attribute (M). We distinguished between items involving spatial order that presumably make (h)igh demands on the correspondence finding process, and other items that make (l)ow demands due to the higher salience of their material attributes. Thus, the component M has two possible attributes: M = {*H*,*L*}, and we define the following diffulty order on these attributes: $H > L$.

A complete item taxonomy

Using the notation we have just defined, it is possible to classify all APM items according to the number and kind of rules involved and according to the difficulty of the material attribute. The main steps in the analysis of the items can be demonstrated with the help of the problem in Figure 1.

For the solution of this item, two rules have to be detected (N={*2*}). Considering only the rows first, they are:

* In each row, the shading of the geometrical figure remains constant („constant in a row“).

* In each row, the three geometrical figures are permutated („distribution of three values“)

Likewise, the columns can be considered. In this case, the rules are:

* In each column, the intensity of the shading is growing („pairwise progression“).

* In each column, the three geometrical figures are permutated („distribution of three values“).

Thus, the following demand components can be formulated: A first rule operates on the material component „geometrical figure“. Considering the row, this rule is „distribution of three values“ (D3). The same rule applies columnwise. A second role governs the pattern. Considering the row, this rule is „constant in a row“ (CR). Considering the column, it is „pairwise progression“ (PP). Carpenter et al. (1990) state that in their experiments most subjects analyzed the problems by rows. However, there is no way of being sure that all subjects will always analyze the problems by rows. To assure that all subjects have to find the same solution rules regardless of whether they employ a rowwise or columnwise analysis, it is important that items are governed by the same rules columnwise and rowwise. Otherwise, the claim can not be made that the test is the same for all subjects. Fortunately, in the vast majority of problems the rule types are the same regardless of whether a row or column organization is applied (Carpenter et al., 1990). Moreover, we found that among the remaining cases (one of which is depicted in Figure 1) there are only two variants in which alternative prerequisites concerning the detection of a rule occur columnwise and rowwise. These two variants are: pairwise progression or distribution of three values in one direction and constant in a row in the other direction. Based on observations of Carpenter et al. (1990), it is plausible that in these two cases, the most simple and salient operation „constant in a row“ will be decisive for the characterization of item difficulty. With regard to the distinction between rules that do or do not involve an exclusive-or conjunction, however, this assumption is not important (T=$\{O\}$). The material attributes in this item are „shading“ and „geometrical figures“; the more difficult material attribute „spatial order“ does not play a role (M=$\{L\}$).

On the basis of the demand components that were defined above, the analysis of the item depicted in Fig. 1 showed that there are two rules involved (N =$\{2\}$); that there are only types of rules other than the exclusive-or conjunction involved (T=$\{O\}$); and that the difficulty of the material attributes is low (M=$\{L\}$) because spatial order is not important. The demand components of this example can therefore be written as (*2,O,L*). In a similar way, an unequivocal description of the demand components of all other items is possible. Our analyses showed, however, that due to a lack of suitable distractors a solution for some of the items (item 1, 3, 21 and 25 of set II) can already be found on the basis of a subset of the rules that are contained in the respective item after eliminating all distractors but one (cf. Vodegel Matzen, 1994). Because there is evidence that especially when trying to solve the more difficult items, participants often adopt an elimination strategy instead of deducing all existing rules (Carpenter et al., 1990), we decided to consider only the prerequisites that are minimal for a successful solution of each item. We will return to the question of how to construct more suitable distractor items on theoretical grounds in the discussion section.

- Table 1 about here -

Table 1 summarizes the number of rules involved and the frequency of occurence of different rules and material attributes in the 46 items of the APM. For example, there are ten items in which three rules are involved; the rule „constant in a row“ shows up once in 6 items, and it shows up twice in another 11 items. A list of all items sorted by their respective demand components as defined above is shown in Table 2. The complete list of the demand components we identified for the 46 items of the APM is given in Appendix A. This appendix also shows the rules contained in each item and the proportion of subjects that solved each item in a large reference sample of 1015 Germans of both sexes that was examined by Kratzmeier (1976).

- Table 2 about here -

Methods

Having identified the components which influence item difficulty, the most important prerequisite to establish a formal structure over the problem components is met. The Theory of Knowledge Spaces introduced by Doignon and Falmagne (1985) offers a useful framework for this undertaking. Since we deal with the performance of the participants of an intelligence test, in this study we prefer to use a slightly different notation and speak of *performance spaces*, following Korossy (1996). However, the formal theory and its interpretation is the same.

A formal theory of performance

We adopt the following definitions introduced by Doignon and Falmagne (1985) and discussed by Falmagne, Koppen, Villano, Doignon and Johannesen (1990):

The domain of a test is conceptualized as a set Q of items. The performance of an individual with respect to that domain is characterized as a subset $K \subseteq Q$ of all the problems that this individual is capable of solving. K is called the subject's solution pattern. The family of all possible solution patterns is obviously the power-set 2^Q. It is reasonable to assume, however, that not every possible subset of Q is also a reasonable solution pattern. For example, it is plausible that a subject that is able to solve an item consisting of two rules of a certain kind is also able to solve another item that consists of only one rule of the same kind. Defining $K \in 2^{Q}$ as the set of all *reasonable* solution patterns we put some algebraic structure on Q. In other words, we assume that some solution patterns are not *admissible* in a sense that will be defined more explicitly now.

In order to determine all reasonable solution patterns one can try to structure the problem set Q by means of a so-called *surmise relation S* on Q such that for any items $q,r \in Q$ we have the interpretation

(1) $q\ S\ r$ if and only if any subject who is able to solve r should also
be able to solve q.

For any pair (q,r) of items with qSr, item q is called a *prerequisite* of item r. In other words: if a participant is able to solve item r, he or she will also be able to solve item q.

In the case that the binary relation S is a linear ordering of the item difficulties it is strongly related to the well-known notion of a Guttman scale (Guttman, 1947, 1950). In fact, the theory of performance spaces generalizes this concept. Thus, the assumption of a one-dimensional (homogeneous) item structure, which has been violated in many applications of Guttman scaling and its probabilistic version in latent trait theory and which is also questionable on the basis of the above mentioned factor analytic results, is no longer necessary. Rather, S is more generally supposed to be a quasi order (i.e., reflexive and transitive). The notion of a surmise relation S is introduced and extensively discussed by Doignon and Falmagne (1985). For any quasi order on S, the set of all solution patterns □ *compatible with S* can be defined by

(2) $□ := \{ K \subseteq Q \mid \text{if } r \in K \text{ and } qSr, \text{ then } q \in K \text{ for all } q,r \in Q\}$.

According to this definition a solution pattern $K \subseteq$ Q is an element of □ if and only if it contains with any item $r \in K$ also all prerequisites of item r (every item $q \in Q$ with qSr). The solution patterns that agree with the surmise relation are called *performance states.*

Doignon and Falmagne (1985) show that □ as defined above can be axiomatized algebraically. In particular, □ is closed under union and intersection. Conversely, any set of solution patterns closed under union and intersection of sets can be characterized by a uniquely determined surmise relation S on Q according to a theorem of Birkhoff (1973).

One way to represent a quasi order S is a Hasse diagram. In a Hasse diagram, the corresponding surmise relation is shown in a very economical way: lines for „reflexive“ ordered pairs such as (r,r) and for ordered pairs which can be derived through transitivity (e.g., if (s,r), $(t,s) \in S$, then $(t,r) \in S$) are omitted. It is assumed that every person that solves a particular problem also solves all other problems which are below this problem and connected through a line or a downward series of connected lines (for a detailed definition, see Davey & Priestley, 1990, p.7).

An example is shown on the left side of Figure 2. We have a set $Q = \{r,s,t\}$ of problems on which a quasi-order $\{(r,r),(s,s),(t,t),(s,r),(t,r)\}$ is defined. Relation S is shown as a Hasse diagram. The diagram shows that the solution of items s and t can be inferred from the solution of item r. It is assumed, however, that problems s and t cannot be compared in this manner: since there is no downward line between the two items, from the fact that a participant solves item s nothing can be said about whether he also will solve item t, and vice versa.

Another example is shown on the right side of Figure 2. On a set $Q = \{r,s,t\}$ of problems a linear order $\{(r,r),(s,s),(t,t),(s,r),(t,s),(t,r)\}$ is defined. From the Hasse diagram we can see that this order is a special case of a quasi-order, because every problem is comparable to all other problems. Problem r, for example, is supposed to be more „difficult“ than problems s and t. Problem s is supposed to be more difficult than problem t. This type of problem ordering is known as a Guttman scale (Guttman, 1947, 1950).

- Figure 2 about here -

Several approaches can be used to determine □, the set of all admissible solution patterns. One possibility is to query experts about the likely relationships between different problems of a given domain (Koppen & Doignon, 1990; Dowling, 1993). Another possibility is the analysis of solution patterns in large empirical data sets (Falmagne, 1989; Villano, 1991). A third, alternative approach to determine □ developed by Albert and Held (1994) is to formulate structural relationships between test items based on a thorough task analysis. This latter approach appears to be useful for the formulation of a mathematical theory of performance on the APM: surmise relations are established on the basis of basic problem elements of each item which are called *problem components.* Generally, a component is associated with a *demand* of the problem. We have already ordered the attributes of the three components N, T and M according to their difficulty. Using these orderings on the attributes of each component, it is possible to derive a surmise relation for the APM items. Two methods come into question for this undertaking: the *componentwise ordering of products* and the *lexicographic ordering of components* (Albert & Held, 1994).

A *componentwise ordering of products* can be established between the components of a problem by defining each component as a set of *attributes,* which can be substituted for the component. Problems are generated by forming the Cartesian product of all considered components. For the establishment of a problem structure (surmise relation), all problems q_x, q_y have to be compared pairwise, with respect to the attributes of the components. Formally, the ordering rule is defined as follows:

Let $P_1, \ldots, P_n$ be component sets on which partial orders $R_1, \ldots, R_n$ are defined. On the Cartesian product $P_1 \text{ x } \ldots \text{ x } P_n$, which corresponds to the possible item set, an order $\leq$ is imposed by defining

(3) $(p_1, .. , p_n) \leq (q_1, .., q_n)$ iff for all i: $p_i\ R_i\ q_i$ with $p_i, q_i \in P_i$.

Expressed in words: we surmise that a problem q is at least as difficult to solve than a problem p, if all attributes $q_1, .., q_n$ of q are at least as difficult as the corresponding attributes in p with respect to the relations defined on the attribute sets. This principle is known as „coordinatewise order“ (for a description see Davey & Priestley, 1990, p. 19). Note that this method is also known from decision theory where the choice heuristic called *dominance rule* corresponds to the coordinatewise order. According to Birkhoff (1973), $\leq$ is a partial order.

Since the attributes of the components must be compared, it is necessary to define an order on each set of attributes.

A simple example, namely, the dominance rule based model D(NxM), might help to clarify the principle of a dominance rule. The model considers two components, the number of rules N = {*1,2*} and the difficulty of the material attribute M = {*H,L*}. According to the dominance rule, an item (*2,H*) is expected to be more difficult than the items (*1,H*), (*2,L*), and (*1,L*) because its attributes are more difficult than the corresponding attributes of these other items with respect to at least one of the components. However, there is no such relationship between the items (*1,H*) and (*2,L*): the former item is more difficult with respect to its material attribute, but the latter item is more difficult with respect to the number of rules involved.

- Figure 3 about here -

The *lexicographic ordering of components* can be viewed as a special case of componentwise ordering. Aside from component orders, an order of „importance“ defined on the set of all components is assumed. According to this, the attributes of the most important component are compared first. If this ordered pair of attributes is an element of the respective component order on the attributes, the problem with the more difficult attribute is supposed to be more difficult as a whole (i.e., the problem pair is an element of the surmise relation). Otherwise, the second most important component is focussed. This procedure is known from dictionaries, where words are ordered this way (beginning with the first letter), and continues on until two different attributes are found for the *n*th component or until there are no more components left to compare. Formally, the imposed order always is linear (transitive, anti-symmetric, and connected) and thus comparable to a Guttman scale (see Albert & Held, 1994, for a more detailed discussion). Lexicograpic orders impose very strict constraints on the possible relationships between a set of items. If only one component is considered, however, the dominance model and the lexicographic model cannot be distinguished.

Again, a simple example, namely, the model L(N>M), might help to illustrate the concept of a lexicographic rule. Let us consider the same components as before, the number of rules N = {*1,2*} and the difficulty of the material attribute M = {*H,L*}. This time, we assume that the difficulty of the material attribute subordinates to the number of rules (N>M). The resulting surmise relation therefore predominantly reflects the number of rules involved, and only if two items involve an equal number of rules, the difficulty of their material attribute decides about their relative order.

- Figure 4 about here -

Of course, the question of whether a dominance rule or a lexicographic rule is better suited to describe the relationship between test items is an empirical one. In our opinion, the question of which components have to be taken into consideration for an appropriate modelling of item difficulty should also be decided empirically. Therefore, we decided to compile a list of all models that come into question. The models are either based on a dominance rule or a lexicographic rule between all problem components or a subset of the problem components described above. The resulting list of models is shown in Table 3.

- Table 3 about here -

All resulting problem orderings identified by the seven models with a dominance rule and the twelve models with a lexicographic rule can be represented as Hasse diagrams. The further proceeding will be illustrated with the help of model D(N x T). It is represented with the help of the Hasse diagram shown in Figure 5.

- Figure 5 about here -

The Hasse diagram of the model D(NxT) visualizes the problem ordering that characterizes this model. It is assumed that every person that solves a particular problem also solves all other problems that are located below the problem and are connected through a line or a downward series of connected lines. For example, a person who is able to solve problem (*3,O*) should also be able to solve problems (*2,O*) and (*1,O*). The dominance rule, however, does not make a prediction for (*3,O*) concerning the problems (*1,X*) and (*2,X*). Compared to the problem (*3,O*), these problems are characterized by a lower number of rules, but their material attribute is more difficult. For each problem type the number of items representing the respective problem type has been counted using the item taxonomy in Appendix A (cf. also Table 2). It is important to note that among the 46 items of the APM there is no problem of the type (*4,X*). On the other hand, there are 12 problems of the type (*1,O*), reflecting a very unbalanced distribution of items representing the different problem types. According to the normative data of Kratzmeier (1976), 80 percent of the subjects solve the 12 problems of type (*1,0*), a number that is depicted in italic letters in Figure 5. There is exactly one item of the type (*3,X*). It was solved by only 4 percent of the reference population. The fact that the number of items representing the different types of problems is not balanced has consequences for the analysis of individual solution patterns that is performed later on. We will return to this issue later.

Model selection based on aggregate data

The solution percentages according to Kratzmeier (1976) make a first rough screening of our models possible. For this screening, the following criterion was used: a model is violated if a problem type with higher difficulty according to the model is solved by more persons than a problem with lower difficulty according to the model. As can be seen, no problem violates this criterion in the case of model D (N x T) in Figure 5 (80 > 68 > 41 > 24, 47 > 27 > 4, 80 > 47, 68 > 27, and 41 > 4).

The chosen criterion - item difficulty may not be greater for a problem that is, according to the respective model, easier than another problem - was adopted to all dominance-rule based and all lexicographic-rule models.

Of these models, only four models satisfy the criterion. Among the dominance-rule models, these are the models D(N), D(NxT) - cf. Figure 5 - and D(TxM). The least restrictive dominance-rule model D(NxTxM) and three other dominance-rule based models in Table 2 are not compatible with the criterion. Most of the lexicographic-rule models are not compatible with the criterion either, with the only exception of the model L(T>M) that does not take into account the number of rules, but rather assumes that type of problem (T) and, subordinate to this component, the difficulty of the material attribute (M) are the most important determinants of item difficulty.

The Hasse diagram for model D(N x T) was already presented in Figure 5. The Hasse diagram for model D(N) is shown in Figure 6; Figure 7 shows model D(TxM), and model L(T>M) is shown in Figure 8.

- Figure 6 about here -

The model D(N) classifies all problems solely according to the number of rules involved. The resulting order is equivalent to a Guttman scale. Because no other components are considered in this model, it is also identical to the model L(N).

The model D(TxM) depicted in Figure 7 classifies the problems according to the type of rule and the difficulty of the material attribute. It does not consider the number of rules involved (N). Regrettably, there is no item in the APM that represents the problem type (X,H).

- Figure 7 about here -

The more specific model L(T>M) assumes that in the first order, the type of rule determines item difficulty (Figure 8). According to this model, items involving an exclusive-or rule are assumed to be more difficult than items without this rule, regardless of the difficulty of the material attribute. Only if both items are comparable with respect to rule difficulty, the difficulty of the material attribute decides on the order of the items. Again, the resulting order is equivalent to a Guttman scale on these problems. This model is the only model based on a lexicographic rule that is compatible with the screening criterion (the model L(N) is equivalent to the dominance model D(N)). As noted before, there is no item in the APM that represents the most difficult problem type according to the model: (X,H).

The results for the lexicographic models indicate that none of the components is subordinate to another with the exception of the material attribute (M), which is subordinate to the type of rule (T) with respect to its contribution to item difficulty. Therefore, L(T>M) was retained for a more detailed analysis. However, it is not possible to subordinate any component to the number of rules involved (N) as indicated by a lack of fit of the models L(N>T), L(N>M), L(N>M>T) and L(N>T>M). Nor subordinates another component the number of rules involved

(N) as indicated by the lack of fit of the models L(T>N) and L(M>N). All of these models violate the criterion in one or several positions of the respective Hasse diagrams; it follows that the even more restrictive models L(T > N > M), L(T > M > N), L(M > N > T) and L(M > T > N) also fail to meet the criterion.

- Figure 8 about here -

To summarize, four models satisfy the chosen criterion: D(N), D(NxT), D(TxM) and L(T>M). The least restrictive dominance rule based model D(NxTxM) which is presented in Figure 9 is not compatible with the criterion. However, it is violated only because of the problem *(3,O,H)* which easier than expected (49% solvers). A percentage of solvers between 21 and 35 on this item would have led to the retention of the model (cf. Figure 9). As can be seen from Appendix A, in two of the four items representing this problem type (Item I/9 and II/13) the difficult material attribute (H) is paired with the most simple rule „constant in a row“ (CR). As indicated by the high solution rates of 64% and 61%, respectively, this might have led to an unusually easy processing of the difficult material attribute due to the application of a non-analytic perceptual „Gestalt algorithm“ (Hunt, 1974). For this reason it was decided to retain the model (DxTxM) for subsequent analysis despite the small violation of the model according to the group data criterion.

- Figure 9 about here -

A cross-validation with individual answer patterns

Having identified five candidate models, it is necessary to leave the level of group data and cross-validate the previous results with the help of individual answer patterns. This cross validation was conducted in order to determine which of the five models that best satisfy the group level criterion are also suited for the description of individual answer patterns.

One of two data sets used for the cross validation was obtain by administering the APM to a sample of undergraduate students at the University of Graz[1] (n = 56). The participants dealt with a computerized version (Schuhfried, 1987) of both sets of the APM (which was shown to lead to results comparable to the paper and pencil version by Neubauer et al., 1991) within a time limit of 50 minutes, some of them in partial fulfilment of a course requirement. Following the procedure of Alderton and Larson (1990), 8 subjects that skipped one or more items or failed to complete the test in the allotted time were eliminated. This elimination was necessary in order to avoid erroneous conclusions if due to lack of time a subject fails to solve an item that otherwise would have been within his or her capacity. Another 7 subjects wre excluded because they were not cooperative due to lack of time or due to self-reported physical exhaustion in connection with the consumption of alcohol. Thus, the first data set consisted of 41 participants.

A second data set was compiled from two previously published investigations of Neubauer (1990b; Neubauer, Urban & Malle, 1991)[2]. From the study of Neubauer (1990b), those undergraduate students who attempted to solve all items of set II within a time limit of 40 minutes (no time limit was set for the items of set I) were selected, resulting in a sample of 14 students. In another study of Neubauer et al. (1991), one part of the sample dealt with the paper and pencil version of the APM; additional participants that were first administered a computerized version of the APM before solving the paper and pencil version were not be included in our analysis because of significant practice and learning effects. From the paper and pencil sample, those 30 participants who tackled all items of set I within the time limit of 10 minutes and all items of set II within the time limit of 40 minutes were selected. Thus, the individual answer patterns of 44 undergraduate students of both sexes were included in the second data set.

We used these two sets of individual answer patterns[3] for a more rigorous test of the models selected above on the basis of aggregate data. In doing so, the above mentioned problem arose that the number of items in the APM representing the different types of problems varies. Due to the possibility of careless errors, the assumption does not seem justified that *all* items of a given problem type would be solved by a person that is capable of solving this problem type in principle. Taking into account the number of 46 items in the test and a probability of a careless error of say only 5%, a criterion as strict as this would result in a probability of only $(1 - 0.05)^{46} = 0.09$ that an observed response pattern really conincides with a participant’s actual ability. It therefore makes sense to define a threshold, the passing of which is assumed to reflect a subject’s ability in principle to solve a given problem type. In order to control for the manipulation of this threshold, we decided to vary it in three steps: 50%, 75%, and 90%. For example, applying a threshold of 75%, at least 3 items out of 4 have to be solved in order to reach the conclusion that a participant is capable of solving the items of this problem type. If less than 4 items (or, respectively, 10 items in the case of a 90% threshold) represent a given problem type, the same conclusion is drawn only if all items representing the respective problem type are solved without any exception. We felt that these thresholds are high enough to prevent erroneous conclusions due to (a series of) lucky guesses, which have an a priori probability of 1/8=12.5% according to the number of distractors.

Results

The fit of all of the considered models to the individual answer patterns from two data sets are shown in Table 4. For each of the different threshold values, the *symmetrical distance*[4] between the different performance spaces and the individual answer patterns was determined (cf. Albert, Schrepp & Held, 1994). *Mean symmetrical distance* denotes the average number of deviations between each person’s response pattern and the closest admissible performance state according to the respective model: the smaller the symmetrical distances are, the better is the fit of the respective model. Table 4 also provides the number of states and nonstates (admissible and forbidden answer patterns) for each model. Nonstates are solution patterns (subsets of the problem set) which do not belong to („are not admissible“ according to) the performance space of a given model. States are possible solution patterns that are in accordance with the respective model. Table 4 also shows the number of *different* congruent and noncongruent response patterns that were observed empirically. This is important in order to judge whether a sizeable proportion of the hypothesized performance states indeed could be observed empirically, and for detecting possible floor or ceiling effects. Also shown is the frequency distribution of the symmetrical distances between response patterns and states, both for the two empirical data sets and for the average of the 1000 simulated random patterns that were generated for each model by determining with probability .5 whether each item was solved or not[5]. One indicator for the goodness of fit of a model is a large number of zero distances indicating solution patterns that are predicted by and are therefore in accordance with the respective model. Furthermore, from the distance distribution, it is possible to calculate the average symmetrical distance for each model. It is important to note that for the more complex models with a larger proportion of nonstates, the symmetrical distance will always be greater by chance alone. Because of the large variation in the number of nonstates which are hypothesized in the different models, the average symmetrical distance of the two empirical data sets is therefore compared to the average symmetrical distance of simulated random response patterns. The resulting *coefficient of distance agreement* (DA coefficient; Schrepp, 1993) is calculated as the quotient

$$\text{DA} \quad := \quad \frac{\text{mean average distance of an empirical data set}}{\text{mean average distance of random response patterns}}.$$

It can be used as a measure for the goodness of fit in order to discriminate between models that postulate performance spaces of different sizes. The DA coefficient assumes a value of 0 if the fit of a model is perfect; values about 1 indicate a bad fit that is not better than chance (cf. Schrepp, 1993).

One first essential result of the model tests on an individual basis is that all five of the considered performance spaces display a very high goodness of fit according to conventional criteria. The chi-square for the comparison of the distances of random answer patterns and the two sets of empirical answer patterns is significant ($p < .01$) regardless of the threshold chosen. As can be seen in Table 4, there is a high number of zero or unit distances both in absolute and relative terms, indicating a perfect or near perfect congruence between observed answer patterns and admissible states according to the respective models. This finding replicates in both samples.

From the distributions of distances, the mean symmetrical distance was computed, which in turn was used to determine the distance agreement (DA) coefficient relative to the average distance of response patterns. As can be seen, the mean symmetrical distance is below 1.0 in all cases, also indicating a very good fit of the models.

A second result is that the distance measures are not very sensitive to changes in the threshold chosen to determine the percentage of items that have to be solved in order to conclude that a participant is capable of solving the items of a given problem type. The conclusion seems justified that the choice of threshold is not an important determinant of goodness of fit, strengthening the confidence in the results of our study.

- Table 4 about here -

The DA coefficient offers a stricter criterion that can be used in order to differentiate between the five models that appear to be comparable on the basis of conventional chi-square criteria. On the basis of the DA coefficients, the models D(N) and D(TxM) should be preferred over the models D(NxT), D(NxTxM) and L(T>M). The fit of the two former models is excellent, regardless of which sample is considered and regardless of the threshold chosen. By far the most observed answer patterns are in perfect accordance with these two models. The DA-coefficients of the two models D(N) and D(TxM) vary in the range of .00 to .26; the mean symmetrical distances vary between .00 and .16. Since all hypothesized performance states can be observed empirically, the high goodness of fit is clearly not due to possible floor or ceiling effects in the participant's performance. The models D(NxT) and D(NxTxM) have slightly higher, but still satisfying mean symmetrical distances and distance agreement coefficients.

The results show further that L(T>M), the only model with a lexicographic rule that seemed applicable in view of the aggregate data of Kratzmeier (1976) does not always lead to a satisfying goodness of fit if applied to individual answer patterns, as indicated by DA coefficients varying between .05 and an unsatisfying .68 depending on the threshold chosen. Results thus indicate that dominance rule based models are better suited than lexicographic rule based models to describe the relationship between the components number of rules (N), type of rule (T) and difficulty of the material (M) in the items of the APM.

Discussion

Michel (1964) suggested that all tests should be conceived as experiments, with the items being the independent variables that are set by the experimenter in order to assess their effect on the test participants. There is no question that theoretical knowledge and control of the independent variable is central in any experiment. Why, then, should test items be constructed free-handed and declared as indicative of a certain aspect of ability only after the theory-free trial-and-error procedure of empirical validation, as is often the case in classical test theory? Item construction has to be more than a posteriori item selection (Lienert, 1969; Thorndike, 1982). Rather, it seems worthwile to look for a theory that allows from the very beginning to construct items of prespecified properties and difficulty. In contrast to classical test theory which does not make any assumptions about the factors that determine item difficulty, we argue that modern test construction ought to be based on a theory of the ability concept in question. Items then need not be justified by means of *a posteriori* statistical analyses any longer, but can be constructed following construction rules derived from substantial theory. This will not only contribute to measurement precision and validity, but to interpretability of test scores as well.

The original items of the APM were designed on intuition as to what constitutes item difficulty. Accordingly, the only information that is gained when administering the test is the test sum score, which can be compared to normative group data. We tried to show that a formal theory linking item properties and observable solution patterns provides the basis for a more fine-grained diagnosis of the participant's ability. We demonstrated the applicability of the Theory of Knowledge Spaces (Doignon & Falmagne, 1985) in the formulation of Albert and Held (1994) for the analysis of the Advanced Progressive Matrices of Raven (1962). We made explicit and falsifiable assumptions about the abilities that are necessary for item solution, establishing theoretically based and formally structured performance spaces which can be used for a diagnostically informative classification of a participant's performance.

The exemplary application of the theory was successful in demonstrating the inadequacy of almost all lexicographic orderings on the components number of rules (N), type of rules (T) and material attributes (M) which were identified as main determinants of item difficulty in earlier studies (e.g., Carpenter et al., 1990; Vodegel Matzen, 1994). Clearly, none of these components is subordinate to another with respect to its influence on item difficulty. The only exception might be the model L(TxM) which does not consider the number of rules but rather imposes a linear order on the items as a function of rule type and difficulty of material. However, this model profits from the fact that there are no items of type *(X,H)* represented in the original form of the APM (cf. Table 2). The inclusion of this problem type in the construction of a new and better balanced set of items would allow a better discrimination between the models L(TxM) and D(TxM).

Dominance rule based models resulting in a partial order on the problems proved to be much more suitable for modelling individual differences in APM performance. In the present investigation, the most simple model D(N) - which is equivalent to model L(N) and considers only the number of rules involved - provided a very good fit to the data. This is in accordance to the proposed unidimensionality of the APM as a pure measure of *g* (cf. Alderton & Larson, 1990). However, to conclude that the APM are unidimensional is certainly premature as becomes evident from inspection of Table 2. As can be seen, the vast majority of the 46 APM items are of type *(1,O,L)*, *(2,O,L)* and *(3,O,L)* (n = 12, 13, and 5, respectively). The low number of items including a) more than two solution rules, b) the more difficult X-or rule and c) the more difficult material attribute in the original version of the APM (cf. Table 2) is a major impediment to a thorough test of the multidimensional models in the present investigation. The unbalanced distribution can easily lead to a deceptive impression of unidimensionality and might also explain why factor analytic investigations have often found one major and many minor factors of item difficulty (Alderton & Larson, 1990). It is our prediction that a better balance will underline the superiority of a multidimensional perspective and allow a much more differentiated evaluation of the models D(NxT), D(TxM) and D(NxTxM).

Another criterion that can be applied to the assessment of performance spaces is the complexity, and, accordingly, the level of detail of a model. This perspective is not very favourable for the model D(TxM) that postulates only four different types of problems with eight different possible solution patterns (because one of the problem types is not represented in the 46 items of the APM). Moreover, no less than five of these eight solution patterns (= 62.5 %) are admissible according to the model. The models D(NxT) and D(NxTxM) clearly are more informative models from a diagnostic perspective, having 128 (1024) possible answer patterns of which only 14 (37) are admissible according to the model. At the relatively low cost of slightly reduced symmetrical distances and coefficients of distance agreement, these models offer the important advantage of a much higher diagnosticity - and falsifiability - than the more simple models D(N) and D(TxM). Further investigations are needed in order to differentiate between these models.

The main problem that became evident during our analysis concerns the fact that there are large differences in the number of items that represent different problem types. Several models suffer from the fact that the more difficult problem types according to the respective model are only rarely and sometimes not at all represented in the APM items (cf. Table 1 and 2). The resulting inaccuracy in determining correct versus wrong responses on an individual basis could be avoided through a better balance in the distribution of the number of items in the different problem types. Moreover, this would also remove unwanted contingencies and associations between demand components in the original set of the APM items that become evident in Table 2. From the theoretical perspective we propose, it seems highly desirable to construct a new set of items on the basis of a componential analysis which provides a better balance among the different problem types. This will also allow a better discrimination among the different models that we investigated.

A major advantage of the present approach is the possibility to construct economical procedures for the adaptive assessment of a person's competence. For an estimation of the solution pattern (the „ability") that characterizes a person, only a portion of the problems needs to be presented. The procedure is similar to that of a good teacher: if one of his questions is answered correctly by a student, the next question will probably be more difficult because the teacher assumes that the student is capable of answering all easier questions, too. The „half-split procedure" suggested by Falmagne and Doignon (1988) guarantees an efficient diagnosis with a minimum number of problem presentations, using a selection rule that is used to halve the number of candidates for a person's competence state by each successive presentation. Selecting always the diagnostically most informative item results in an efficiency gain for n problems of up to $\log_2$ n (in the case of a linear order of problems)[9].

Another problem concerns the choice of distractor items. Extending the idea of a test as a controlled experiment, distractor items can be regarded as an integral part of the stimulus situation and an important determinant of item difficulty. It seems plausible that even a carefully constructed item will not correspond to any definite level of difficulty when the construction of distractors is haphazard. However, we found that the solutions for some of the items can already be found on the basis of a subset of the rules that are contained in these items by eliminating all distractors but one. We therefore decided to consider only the prerequisites that are minimal for a successful solution of each item because there is evidence that especially when trying to solve the more difficult items, less able participants sometimes adopt an elimination strategy instead of deducing all existing rules (Putz-Osterloh, 1981; Carpenter et al., 1990).

There is another point that can be made in the context of distractor items. If only the correct choices are taken into account when measuring performance, no information is extracted from incorrect choices. However, wrong choices can be a good indicator for the misunderstandings of individual subjects (Siegler, 1983). In fact, there have been post-hoc attempts to obtain information from the incorrect choices made by subjects in the APM (Raven, Court & Raven, 1983; Vodegel Matzen, 1994). These analyses were concerned with the identification of different error types as they are represented in the distractors. Vodegel Matzen (1994) constructed a variant of the APM in which all items are based on a priori definitions of the number and kind of rules involved and the errors represented in each distractor. A similar suggestion within the framework of the Theory of Knowledge Spaces has been made by Lukas (1990).

As a consequence of our observations, we suggest that distractors should no longer be constructed on the sole basis of intuition as to what answers might be attractive. Indeed, as White and Zamarelli (1981) have shown, in many tests that were constructed this way so-called „convergent principles“ allow the elimination of most or all distractors prior to any careful analysis of the elements of an item. Through a more systematic construction of distractor items, however, it should be possible to

- avoid the solution of items based on the inspection of the distractor items only,
- make better predictions about the relative empirical difficulties of both test items and distractor items,
- gain additional information from wrong solutions through the analysis of the kind of distractor chosen, revealing the type of error as well as possible partial solutions and misconceptions,
- classify participants according to the types of errors they commit,
- utilize the information gained from choice of distractors for diagnostic purposes, and to
- reduce the variation in test results due to undesired factors („noise“).

Principles according to which distractors can be constructed systematically so as to provide a maximum of diagnostic information were developed by Guttman and Schlesinger (1967) and Hornke and Habon (1984a).

One problem of our approach should be mentioned. Applying discrete models to empirical data always faces an interpretational difficulty: deviations from the predicted states can either be a consequence of wrong model assumptions, or a consequence of some probabilistic mechanism. For example, participants may solve some problems by chance although they do not really have the skills to solve them („lucky guesses“). It is also possible that subjects miss a correct solution to a problem although in principle, they have the skills to solve that particular problem („careless errors“). In such cases, the observed empirical solution pattern does not reflect the participant’s true ability. If the probabilities of such deviations are substantial, a probabilistic model is required. Regrettably, up to now, only preliminary steps towards such a probabilistic model have been taken (Falmagne, 1994). However, representing each problem type with several items and defining an appropriate threshold in order to determine whether a subject is capable of solving a given problem type, as is illustrated in the present study, might provide a feasible solution to this problem.

One last remark should be made. As became evident, the level of abstraction of even the most difficult rules in the APM does not seem particularly high compared with the abstractions that are taught and acquired in various academic domains, such as physics or political science (Carpenter et al., 1990). Jacobs and Vandeventer (1971, 1972) observed training effects in terms of the solution of relations contained on a posttest item over and above material-specific practice effects. The mean training effect obtained by children within an hour or less of individualized instruction represented 15-24 months of „normal growth“. Although one may doubt whether such increases in test scores will result in a concomitant increase in external criteria, it might be possible that further research will result in the creation of training programs for inductive reasoning that lead to improvements in an even broader range of cognitive functions. Indeed, there is some preliminary evidence that well-defined training procedures for inductive reasoning might produce transfer to nontrained contexts in scholastic achievement (Klauer, 1993, 1994) and complex problem solving (Klauer, 1996).

References

Albert, D., & Held, T. (1994). Establishing knowledge spaces by systematical problem construction. In D. Albert (Ed.), *Knowledge structures* (pp. 81-115). Heidelberg, New York: Springer.

Albert, D., & Lukas, J. (1999). *Knowledge spaces: Theories, empirical research, and applications.* Mahwah, NJ: Lawrence Erlbaum.

Albert, D., Schrepp, M., & Held, T. (1994). Construction of knowledge spaces for problem solving in chess. In G. Fisher & D. Laming (Eds.), *Contributions to Mathematical Psychology, Psychometrics, and Methodology* (pp. 123-135). Berlin: Springer.

Alderton, D.L., & Larson, G.E. (1990). Dimensionality of Raven's Advanced Progressive Matrices items. *Educational and Psychological Measurement, 50,* 887-900.

Arthur, W., Jr., Barrett, G.V., & Doverspike, D. (1990). Validation of an information-processing-based test battery for the prediction of handling accidents among petroleum-transport drivers. *Journal of Applied Psychology, 75,* 621-628.

Arthur, W,. Jr., & Woehr, D. (1993). A confirmatory factor analytic study examining the dimensionality of the Raven's advanced progressive matrices. *Educational and Psychological Measurement, 53,* 471-478.

Birkhoff, G. (1973). *Lattice theory* (3rd ed). Providence: American Mathematical Society.

Boring, E.G. (1923). Intelligence as the tests test it. *The New Republic, 34,* 35-36.

Carpenter, P.A., Just, M.A., & Shell, P. (1990). What one intelligence tests measures: A theoretical account of the processing in the Raven progressive matrices test. *Psychological Review, 97,* 404-431.

Cattell, R.B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educationl Psychology, 5(4),* 1-22.

Cattell, R.B. (1971). *Abilities: Their structure, growth, and action.* Boston: Houghton Mifflin.

Court, J.H. (1991). Asian applications of Raven's Progressive Matrices. *Psychologia. An International Journal of Psychology in the Orient, 34,* 75-85.

Court, J.H., & Raven, J. (1982). *Manual for Raven´s progressive matrices and vocabulary scales.* (Research Supplement No. 2, Pt.3, Section 7).

Davey, B., & Priestley, H.A. (1990). *Introduction to lattices and order.* Cambridge: University Press.

Dillon, R.F., Pohlmann, J.T., & Lohman, D.F. (1981). A factor analysis of Raven's Advanced Progressive Matrices freed of difficulty factors. *Educational and Psychological Measurement, 41,* 1295-1302.

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

Dowling, C.E. (1993). Applying the basis of a knowledge space for controlling the questioning of an expert. *Journal of Mathematical Psychology, 37,* 21-48.

Egan, D.E., & Greeno, J. (1974). Theory of rule induction: Knowledge acquired in concept learning, serial pattern learning, and problem solving (pp. 43-103). In L.W. Gregg (Ed.), *Knowledge and cognition.* Hillsdale, NJ: Erlbaum.

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

Falmagne, J.C. (1994). Finite Markov learning models for knowledge structures. In G. Fisher & D. Laming (Eds.), *Contributions to Mathematical Psychology, Psychometrics, and Methodology.* Berlin: Springer.

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

Falmagne, J.C., Koppen, M., Villano, M., Doignon, J.P., & Johannesen, L. (1990). Introduction to knowledge spaces: how to build, test and search them. *Psychological Review, 97,* 201-224.

Fischer, G.H. (1983). Logistic latent trait models with linear constraints. *Psychometrika, 48,* 3-26.

Fischer, G.H., & Formann, A.K. (1982). Some applications of logistic latent trait models with linear constraints on the parameters. *Applied Psychological Measurement, 6,* 397-416.

Formann, A.K. (1982). Linear Logistic Latent Class Analysis. *Biometrical Journal, 24,* 171-190.

Formann, A.K. (1973). *Die Konstruktion eines neuen Matrizentests und die Untersuchung des Lösungsverhaltens mit Hilfe des linearen logistischen Testmodells.* [The construction of a new matrices test and the investigation of solution behaviour with the help of the linear logistic test model]. Unpublished dissertation, University of Vienna.

Formann, A.K., & Piswanger, K. (1979). *WMT - Wiener Matrizen Test.* Beltz: Weinheim.

Guilford, J.P. (1952). When not to factor analyze. *Psychological Bulletin, 49,* 26-37.

Guttman, L. (1947). A basis for scaling qualitative data. *American Sociological Review, 9,* 139-150.

Guttman, L. (1950). The basis for scalogram analysis. In S.A. Stouffer, L.A. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star, & J.A. Clausen (Eds.), *Measurement and prediction* (Vol. 4, pp.60-90). London: Princeton University Press.

Guttman, L., & Schlesinger, J.M. (1967). Systematic construction of distractors for ability and achievement test items. *Educational and Psychological Measurement, 27,* 569-580.

Haygood, R.C., & Bourne, L.E. (1965). Attribute- and rule-learning aspects of conceptual behavior. *Psychological Review, 72,* 175-195.

Held, T. (1994). *Di.* Unpublished computer program, University of Heidelberg.

Hockemeyer, C. (1994). *Bconstra.* Unpublished computer program, University of Braunschweig.

Hornke, L. (1983). Computerunterstütztes Testen - Eine bewertende empirische Untersuchung. [Computer aided testing. An evaluative empirical study]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 4,* 232-344.

Hornke, L., & Habon, M. (1984a). Regelgeleitete Konstruktion und Evaluation von nichtverbalen Denkaufgaben. [Rule-governed construction and evaluation of nonverbal test items]. *Wehrpsychologische Untersuchungen, 19,* 1-153.

Hornke, L., & Habon, M. (1984b). Erfahrungen zur rationalen Konstruktion von Testaufgaben. [Experiences with rational construction of test items]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 5,* 203-212.

Hunt, E.B. (1974). Quote the Raven? Nevermore! In L.W. Gregg (Ed.), *Knowledge and Cognition* (pp. 129-158). Hillsdale, NJ: Erlbaum.

Jacobs, P.J., & Vandeventer, M. (1971). The learning and transfer of double-classification skills: a replication and extension. *Journal of Experimental Child Psychology, 12,* 240-257.

Jacobs, P.J., & Vandeventer, M. (1972). Evaluating the teaching of intelligence. *Educational and Psychological Measurement, 32,* 235-248.

Jensen, A.R. (1987). The g beyond factor analysis. In R.R. Ronning, J.A. Glover, J.C. Conoley, & J.C. Witt (Eds.), *The influence of cognitive psychology on testing,* pp. 87-142. Hillsdale, NJ: Erlbaum.

Keldermann, H., & Rijkes, C. (1994). Loglinear multidimensional IRT models for polytomously scored items. *Psychometrika, 59,* 149-176.

Kinder, A., & Lachnit, H. (1994). Erwerb und Anwendung logischer Relationen bei mehrdimensionalen Reizen [Acquisition and utilization of logical relations with multidimensional stimuli]. *Zeitschrift für Experimentelle und Angewandte Psychologie, 41,* 173-183.

Klauer, K.J. (1993). Induktives Denken beeinflußt das Rechtschreiblernen. [Inductive thinking influences orthographic achievement]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 25,* 352-365.

Klauer, K.J. (1994). Transferiert der Erwerb von Strategien des induktiven Denkens auf das Erlernen eines schulischen Lehrstoffs? [Can the acquisition of inductive thinking strategies be transferred to the learning of a particular school subject?] *Zeitschrift für Pädagogische Psychologie, 8,* 15-25.

Klauer, K.J. (1996). Begünstigt induktives Denken das Lösen komplexer Probleme? [Does inductive reasoning favor the solution of complex problems?]. *Zeitschrift für Experimentelle Psychologie, 43,* 85-113.

Koppen, M., & Doignon, J.P. (1990), How to build a knowledge space by querying an expert. *Journal of Mathematical Psychology, 34,* 311-331.

Korossy, K. (1996). Kompetenz und Performanz beim Lösen von Geometrie-Aufgaben. [Competence and performance in solving geometry problems]. *Zeitschrift für Experimentelle Psychologie, 43*, 279-318.

Kratzmeier, H. (1976). Raven-Matrizen-Test: Advanced Progressive Matrices. Weinheim: Beltz.

Lachnit, H. (1992). Pavlovian conditioning and cognitive psychology: The case of inductive reasoning. *The German Journal of Psychology, 16,* 273-282.

Lienert, G. A. (1969). *Testaufbau und Testanalyse.* [Test construction and test analysis]. Weinheim: Beltz.

Lord, F.M., & Novick, M.P. (1968). *Statistical theories of mental test scores.* Reading: Addison-Wesley.

Lukas, J. (1990). Wissensdiagnose und Fehlerdiagnose auf der Grundlage von Ordnungsstrukturen über Aufgabenmengen. [Knowledge and error diagnosis based on order structures on task sets]. In D. Frey (Eds.), *Bericht über den 37. Kongreß der Deutschen Gesellschaft für Psychologie in Kiel 1990, 99.* Göttingen: Hogrefe.

Lukas, J., & Albert, D. (1993). Knowledge assessment based on skill assignment and psychological task analysis. In G. Strube & K.E. Wender (Eds.), *The Cognitive Psychology of Knowledge.* North Holland: Elsevier.

Lukas, J., & Unnewehr, J. (1993). *Simulation studies for knowledge assessment procedures.* Unpublished manuscript.

McClelland, J.L., & Rumelhart, D.E. (1988). *Explorations in parallel distributed processing.* Cambridge, MA: MIT Press.

McNemar, Q. (1951). The factors in factoring behavior. *Psychometrika, 16,* 353-359.

Michel, L. (1964). Allgemeine Grundlagen psychometrischer Tests. [General foundations of psychometric tests]. In: Heiss, R. (Ed.), *Handbuch der psychologischen Diagnostik,* Bd. 6. Göttingen: Hogrefe.

Musch, J. (1995). *RandDist.* Unpublished computer program, University of Graz.

Nährer, W. (1980). Zur Analyse von Matrizenaufgaben mit dem linearen logischen Testmodell. [The analysis of matrices items with the linear logistic test model]. *Zeitschrift für Experimentelle und Angewandte Psychologie, 27,* 553-564.

Neisser, U., & Weene, P. (1962). Hierarchies in concept attainment. *Journal of Experimental Psychology, 64,* 640-645.

Neubauer, A.C. (1990a). Coping with novelty and automatization of information processing: An empirical test of Sternberg's two-facet subtheory of intelligence. *Personality and Individual Differences, 11,* 1045-1052.

Neubauer, A.C. (1990b). Speed of information processing in the Hick paradigm and response latencies in a psychometric intelligence test. *Personality and Individual Differences,11,* 147-152.

Neubauer, A.C., Freudenthaler, H., & Pfurtscheller, G. (1995). Intelligence and spatio-temporal patterns of event-related desynchronization (ERD). *Intelligence, 20,* 249-266.

Neubauer, A.C., Urban, E., & Malle, B.F. (1991). Ravens Advanced Progressive Matrices: Computerunterstützte Präsentation versus Standardvorgabe [Raven's Advanced Progressive Matrices: Computer-assisted presentation vs. standard presentation]. *Diagnostica, 37,* 204-212.

Pajarez, F., & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology, 20,* 426-443.

Pellegrino, J.W., & Glaser, R. (1979). Cognitive correlates and components in the analysis of individual differences. *Intelligence, 3,* 187-214.

Posner, M.I., & Mitchell, R. (1967). Chronometric analysis of classification. *Psychological Review, 74,* 392-409.

Putz-Osterloh, W. (1981). *Problemlösungsprozesse und Intelligenzleistung.* [Problem solving processes and intelligence test performance]. Bern: Huber.

Raven, J.C. (1962). *Advanced Progressive Matrices, Set I and II.* London: Lewis.

Raven, J.C., Court, J.H., & Raven, J. (1983). *Manual for Raven's progressive matrices and vocabulary scales: Sec. 4. Advanced progressive matrices.* London: Lewis.

Rhenius, D., & Heydemann, M. (1984). Lautes Denken beim Bearbeiten von RAVEN-Aufgaben. [Thinking aloud while working on problems from Raven's matrices]. *Zeitschrift für Experimentelle und Angewandte Psychologie, 2,* 308-327.

Rosch, E. (1977). Classification of real-world objects: origins and representations in cognition. In P.N. Johnson-Laird & P.C. Wason (Eds.), *Thinking: Readings in Cognitive Science* (pp.212-222). Cambridge, GB: Cambridge University Press.

Scheiblechner, H.H. (1972). Das Lernen und Lösen komplexer Denkaufgaben. [The learning and solving of complex reasoning tasks]. *Zeitschrift für Experimentelle und Angewandte Psychologie, 19,* 476-505.

Schrepp, M. (1993). *Über die Beziehung zwischen kognitiven Prozessen und Wissensräumen beim Problemlösen.* [On the relationship between cognitive processes and knowledge spaces in problem solving]. Doctoral dissertation, University of Heidelberg.

Schuhfried, G. (1987). *Wiener Testsystem II.* [Computer Program]. Mödling, Austria.

Siegler, R.S. (1983). Information processing approach to development. In P.H. Mussen (Ed.), *Handbook of Child Psychology.* (Vol. 1, pp. 129-211). New York: John Wiley.

Snow, R.E., Kyllonen, P.C., & Marshalek, B. (1984). The topography of learning and ability correlations. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol.2). Hillsdale, NJ: Erlbaum.

Sternberg, R.J. (1977). *Intelligence, information processing and analogical reasoning: The componential analysis of human abilites.* Hillsdale, NJ: Erlbaum.

Sternberg, R.J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* New York: Cambridge University Press.

Thorndike, R.L. (1982). *Test theory and test construction. Applied Psychometrics.* Dallas: Houghton & Mufflin.

Unnewehr, J. (1992). *Prozeduren zur Wissensdiagnose (Benutzerhandbuch). [Procedures for the diagnosis of knowledge. User Manual]*. Bericht aus dem Psychologischen Institut der Universität Heidelberg, Nr. 74. [Report No. 74, Psychological Institute of the University of Heidelberg].

Villano, M. (1991). *Computerized knowledge assessment: Building the knowledge structure and calibrating the assessment routine.* Unpublished PhD thesis, New York University.

Vodegel Matzen, L. (1994). *Performance on Raven's Progressive Matrices - what makes a difference?* Unpublished PhD thesis, University of Amsterdam.

Vodegel Matzen, L., van der Molen, M., & Dudink, A. (1994). Error analysis of Raven test performance. *Personality and Individual Differences, 16,* 433-445.

Ward, J., & Fitzpatrick, F. (1973). Characteristics of matrices items. *Perceptual and Motor Skills, 36,* 987-993.

White, A.P., & Zamarelli, J.E. (1981). Convergent principles: Information in the answer sets of some multiple-choice intelligence tests. *Applied Psychological Measurement, 5,* 21-27.

Whitely, S. (1980). Modeling aptitude test validity from cognitive components. *Journal of Ecuational Psychology, 72,* 750-769.

Appendix A

Set	Item	Rule(Attribute)	Number of rules	Demand components	Percentage of successful solvers
I	3	1. PP(L)	1	*(1,O,L)*	88
	4	1. PP(L)	1	*(1,O,L)*	89
	5	1. CR(L) 2. CR(L)	2	*(2,O,L)*	86
	6	1. CR(L) 2. CR(L)	2	*(2,O,L)*	75
	7	1. CR(L) 2. D3(L)	2	*(2,O,L)*	88
	8	1. D3(L) 2. D3(L)	2	*(2,O,L)*	71
	9	1. CR(L) 2. CR(H) 3. D2(L)	3	*(3,O,H)*	63
	10	1. FA(L)	1	*(1,O,L)*	82
	11	1. FA(L) 2. FA(L)	2	*(2,O,L)*	42
	12	1. FA(L)	1	*(1,O,L)*	72
II	1	1. CR(L) 2. D3(L)	2	*(2,O,L)*	85
	2	1. CR(L) 2. CR(L)	2	*(2,O,L)*	84
	3	1. CR(L)	1	*(1,O,L)*	84
	4	1. PP(L)	1	*(1,O,L)*	82
	5	1. CR(L) 2. CR(L)	2	*(2,O,L)*	78
	6	1. CR(L) 2. CR(L)	2	*(2,O,L)*	84
	7	1. FA(L)	1	*(1,O,L)*	78
	8	1. D3(L) 2. D3(L)	2	*(2,O,L)*	74
	9	1. FA(L)	1	*(1,O,L)*	83
	10	1. PP(L)	1	*(1,O,L)*	76
	11	1. FA(L)	1	*(1,O,L)*	76
	12	1. FA(L)	1	*(1,O,L)*	79
	13	1. CR(L) 2. CR(H) 3. D3(L)	3	*(3,O,H)*	61
	14	1. CR(H) 2. CR(H)	2	*(2,O,H)*	67
	15	1. PP(L) 2. FA(L)	2	*(2,O,L)*	62
	16	1. FA(L)	1	*(1,O,L)*	67
	17	1. CR(L) 2. D2(L) 3. D2(L)	3	*(3,O,L)*	61

(Appendix continues on next page)

	18	1. PP(L) 2. PP(L) 3. PP(H)	3	*(3,O,H)*	56
	19	1. FA(L) 2. FS(L)	2	*(2,O,L)*	60
	20	1. FA(L) 2. FS(L)	2	*(2,O,L)*	58
	21	1. CR(L) 2. CR(L) 3. CR(L)	3	*(3,O,L)*	45
	22	1. XO(L)	1	*(1,X,L)*	49
	23	1. XO(L)	1	*(1,X,L)*	44
	24	1. CR(L) 2. CR(L) 3. CR(L) 4. CR(L)	4	*(4,O,L)*	30
	25	1. CR(L) 2. CR(H)	2	*(2,O,H)*	36
	26	1. PP(H) 2. D3(L)	2	*(2,O,H)*	40
	27	1. CR(H) 2. D3(L) 3. D3(L) 4. D3(L)	4	*(4,O,H)*	26
	28	1. D3(L) 2. D3(L) 3. D3(L) 4. D3(L)	4	*(4,O,L)*	26
	29	1. CR(L) 2. CR(H) 3. D3(L) 4. D2(L)	4	*(4,O,H)*	20
	30	1. D3(L) 2. D3(L) 3. D3(L)	3	*(3,O,L)*	27
	31	1. D3(L) 2. D3(L) 3. D3(L)	3	*(3,O,L)*	25
	32	1. CR(L) 2. PP(H) 3. D2(L) 4. D2(L)	4	*(4,O,H)*	17
	33	1. XO(L) 2. XO(L)	2	*(2,X,L)*	27
	34	1. D3(L) 2. D3(L) 3. D2(H)	3	*(3,O,H)*	17
	35	1. PP(L) 2. PP(L) 3. FA(L)	3	*(3,O,L)*	18
	36	1. CR(L) 2. CR(L) 3. XO(L)	3	*(3,X,L)*	4

Rules, attributes (in parentheses), number of rules, demand components and item difficulties as observed in a large reference sample of n = 1015 Germans of both sexes (Kratzmeier, 1976) for the 46 items in Set I and Set II of the APM.

Item 1 and 2 of Set I are for use as practice items. Rules are abbreviated as follows: Constant in a row (CR), pairwise progression (PP), distribution of two values (D2), distribution of three values (D3), figure addition (FA), figure subtraction (FS) and exclusive-or (X0). The difficulty of the material attribute, which is given in parentheses after each rule, is either (H)igh or (L)ow. The number of rules in each item varies from 1 to 4.

Appendix B

Answer patterns and marginal frequencies of 41 Graz students (rows) on the 46 APM items (columns).

1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	0	1	0	1	1	37
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	46
1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	42
1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0	0	1	0	1	0	0	0	0	0	0	1	1	0	33
1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	29
1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	0	0	1	1	1	0	1	1	0	0	1	0	0	1	1	1	0	33
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	44
1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	43
1	1	1	0	1	1	0	1	0	1	1	0	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	0	1	1	0	30
0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1	0	1	1	0	1	0	0	0	1	1	0	0	0	1	1	33
1	1	0	1	1	0	0	0	0	0	1	1	0	1	1	1	0	1	1	1	1	0	0	0	0	1	1	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	19
1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	0	0	1	1	0	0	1	1	0	1	0	0	0	34
1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1	1	1	1	0	1	1	0	1	1	0	0	1	1	1	1	38
1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	0	1	0	40
1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	0	1	1	1	1	0	39
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	0	1	0	1	1	0	1	0	1	0	0	37
1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	0	1	1	1	1	0	1	1	0	1	1	0	0	1	1	1	1	0	1	1	0	1	0	0	0	0	1	1	0	32
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	0	1	1	0	0	0	1	1	1	1	0	1	0	38
1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	1	0	0	1	1	0	39
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0	0	1	0	0	1	0	1	1	0	0	0	1	0	35
1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	1	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	29
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	0	1	1	1	0	39
1	1	1	0	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	22
1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	27
1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	0	0	0	0	1	1	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	21
1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	0	0	0	1	1	1	1	1	1	39
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	0	0	0	0	1	0	1	0	1	1	0	1	1	1	0	36
1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	0	1	0	0	24
1	1	1	1	1	1	0	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24
1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	0	0	0	0	1	0	1	0	36
1	1	1	1	1	0	1	1	0	1	1	1	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24
1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	42
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	1	1	0	1	0	0	0	1	0	0	36
1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	1	1	0	1	1	0	1	0	1	1	0	34
1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0	0	0	1	1	0	0	0	1	0	1	1	0	33
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	1	0	1	0	0	0	1	1	0	1	0	0	1	0	1	0	1	0	33
1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	0	0	1	1	0	0	1	0	0	0	1	1	1	1	0	35
1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	26
1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	1	1	0	0	0	1	0	0	0	1	0	0	31
1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	1	1	1	1	1	0	1	1	0	1	1	0	0	0	0	1	0	0	0	31
1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	1	1	1	1	0	0	0	1	0	0	0	0	35
40	41	38	32	41	37	33	36	24	37	40	39	39	40	36	40	38	37	38	38	40	32	27	37	37	35	31	32	25	30	24	24	20	15	22	25	24	17	9	22	17	14	18	22	26	9	

Appendix C

Answer patterns and marginal frequencies of 44 participants from Neubauer (1990; Neubauer et al., 1991) on the 46 APM items (columns).

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0	42
1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1	41
1	1	1	1	1	1	1	1	0	0	1	1	0	1	1	1	1	1	1	1	0	1	1	0	0	0	1	1	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	24
1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	0	1	0	1	0	1	1	1	1	0	0	1	0	1	0	1	1	0	0	0	1	1	33
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	1	1	0	41
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	1	1	1	0	0	0	1	0	0	1	1	0	1	1	0	0	35
1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	0	0	1	39
1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	29
1	1	1	0	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	25
1	0	1	0	1	0	0	0	1	0	0	1	1	0	1	1	1	0	1	1	1	1	0	0	1	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	18
1	1	0	0	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	0	1	0	1	1	0	1	0	0	1	1	0	1	1	0	1	0	31
1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	0	0	1	0	0	33
1	1	0	0	1	1	0	0	0	0	1	0	1	0	0	0	1	1	0	1	0	0	1	0	1	0	0	0	0	1	0	0	0	1	0	0	0	1	1	0	1	0	0	0	0	0	16
1	1	1	1	0	1	0	0	0	1	1	1	1	0	0	1	0	0	1	1	1	1	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18
1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	0	0	1	39
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	1	1	1	1	0	1	1	0	1	1	0	0	0	0	0	35
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	1	0	0	0	0	0	0	1	1	0	36
1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	1	1	1	0	40
0	1	0	1	1	0	0	0	0	1	1	1	1	0	1	1	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	1	1	0	0	0	0	1	0	0	1	0	0	0	17
1	1	1	1	1	1	0	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	0	1	1	0	0	0	1	0	0	1	1	0	0	1	1	0	32
1	1	1	0	1	1	0	1	0	0	1	1	1	0	1	1	1	1	1	1	1	0	0	0	0	1	0	1	1	1	0	1	1	0	1	0	0	0	0	0	1	0	0	0	0	0	24
1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	1	0	1	1	1	0	37
1	1	1	0	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	39
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	0	1	0	0	1	1	0	0	0	36
1	1	1	1	1	1	1	0	0	0	1	0	1	1	1	0	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23
0	1	1	1	0	1	0	1	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12
0	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	30
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	1	0	0	1	1	0	40
1	1	1	0	1	1	0	0	0	0	1	1	1	1	1	0	1	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	17
1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	1	0	1	0	1	0	39
1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	0	1	0	1	1	0	0	0	0	1	0	1	0	0	31
1	0	1	0	1	1	1	0	0	1	1	1	1	1	1	1	0	1	1	0	1	1	0	0	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0	1	1	0	0	0	1	0	26
1	0	1	0	1	0	0	0	1	0	0	1	1	0	1	1	1	0	1	1	1	1	0	0	1	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	18
1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1	41
1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	0	1	0	1	0	1	1	1	1	0	0	1	0	1	0	1	1	0	0	0	1	1	33
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0	42
1	1	1	1	1	1	1	1	0	0	1	1	0	1	1	1	1	1	1	1	0	1	1	0	0	0	1	1	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	24
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	1	1	0	41
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	1	1	1	0	0	0	1	0	0	1	1	0	1	1	0	0	35
1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	29
1	1	1	0	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	25
1	1	0	0	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	0	1	0	1	1	0	1	0	0	1	1	0	1	1	0	1	0	31
1	1	0	0	1	1	0	0	0	0	1	0	1	0	0	0	1	1	0	1	0	0	1	0	1	0	0	0	0	1	0	0	0	1	0	0	0	1	1	0	1	0	0	0	0	0	16
1	1	1	1	0	1	0	0	0	1	1	1	1	0	0	1	0	0	1	1	1	1	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18
41	41	39	32	38	39	24	34	21	29	41	39	41	34	39	37	40	37	39	38	38	32	27	32	32	32	22	29	32	32	20	23	30	23	11	20	21	15	12	23	30	13	20	15	18	6	

Appendix D

Answer patterns and marginal frequencies of 40 Graz students (rows) on the newly constructed 47 matrix items (columns).

```
    1    0    0    0    1    0    0    0    0    0    0    1    0    0    0    1    0    1    1    0    1    1    1    1    1    1    0    0
0    0    1    0    0    1    0    0    0    1    0    0    0    0    0    0    0    0    0     15
    0    1    0    1    0    1    1    1    0    1    1    1    1    1    1    1    0    1    1    1    1    1    1    1    1    0    0    0
0    1    1    1    1    1    1    1    1    1    0    1    1    0    0    0    1    1    0     33
    0    0    0    0    0    0    0    1    0    0    1    1    0    0    0    1    0    0    1    0    1    1    1    0    1    1    0    1
1    1    1    1    0    0    0    0    0    0    0    0    0    0    0    0    1    0    0     16
    0    1    0    1    0    0    0    0    0    0    1    0    0    0    1    0    0    1    1    1    1    1    1    1    1    0    1    1
1    0    0    0    1    1    1    1    1    0    0    0    1    0    0    0    1    0    0     22
    1    0    1    0    1    1    1    1    1    1    1    0    1    1    1    1    0    1    0    1    1    1    1    1    1    1    1    1
1    1    1    1    0    1    0    1    1    1    1    1    0    1    1    0    0    0    1     36
    1    1    1    1    0    1    1    0    0    0    1    1    1    0    0    1    0    0    0    0    1    1    1    1    1    1    0    1
0    0    1    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    1     19
    0    0    0    0    0    0    1    0    0    0    0    1    1    0    1    0    0    1    1    0    1    1    1    1    1    1    1    1
1    1    0    0    0    1    1    1    1    1    0    1    0    1    1    0    1    0    0     25
    0    0    0    0    1    1    0    0    1    1    1    1    1    0    1    1    0    0    1    1    1    1    1    1    1    1    1    0
0    0    1    1    0    0    0    1    1    0    0    1    0    0    0    0    0    0    0     23
    0    0    1    1    1    0    0    1    1    1    1    1    1    0    0    0    0    0    1    1    1    1    1    1    1    1    0    1
1    0    0    1    0    1    1    0    0    0    0    1    0    0    1    0    0    1    0     25
    0    0    0    0    0    0    0    1    1    0    0    1    1    0    0    0    1    0    1    1    1    1    1    1    1    1    1    1
1    1    0    0    0    0    1    0    1    0    0    0    0    0    0    0    0    1    0     20
    0    1    1    0    0    1    1    1    1    1    1    0    1    0    0    0    0    0    0    1    1    1    1    1    1    1    0    1
1    1    0    1    1    1    1    0    1    0    0    1    0    0    0    0    0    0    0     25
    1    1    0    0    0    0    0    0    0    0    1    1    1    0    0    0    0    1    1    0    1    1    1    0    1    1    1    0
1    1    1    1    1    0    0    0    0    0    1    0    1    0    0    1    1    0    0     22
    0    1    1    1    1    1    0    1    1    0    1    1    1    1    0    1    0    0    1    1    1    1    1    1    1    1    1    1
1    1    1    1    1    1    0    1    1    1    0    0    1    1    1    1    1    0    0     36
    1    0    0    1    1    1    1    0    1    0    1    1    1    0    0    1    0    1    1    1    1    1    1    1    1    1    1    1
1    1    1    1    1    1    1    0    0    1    0    1    0    1    0    0    0    1    0     32
    0    1    0    1    1    0    0    1    0    1    1    1    1    0    0    1    0    1    0    1    1    1    1    1    1    1    0    1
1    1    1    1    0    0    1    0    0    0    0    1    1    0    0    0    0    0    0     25
    1    1    1    1    1    1    1    1    1    1    1    1    1    0    1    1    0    0    1    1    1    1    1    1    1    1    1    1
1    1    0    1    1    1    1    1    1    1    1    1    1    0    1    1    1    0    0     40
    1    1    1    1    1    1    1    1    1    1    1    1    1    0    1    1    1    1    1    1    1    1    1    1    1    1    0    1
0    1    1    1    1    1    1    0    1    1    0    1    1    0    0    1    0    0    0     37
    0    0    1    0    0    0    1    1    1    0    1    1    0    1    1    1    0    1    0    1    1    1    1    1    1    1    1    1
1    0    0    1    0    1    0    1    1    1    0    0    0    1    0    0    0    0    0     26
    0    0    0    0    0    1    0    1    1    1    1    1    1    0    1    0    0    0    0    1    1    1    1    1    1    1    0    1
1    0    1    0    0    1    0    0    1    1    1    0    0    0    0    0    1    1    0     24
    0    1    0    0    0    0    0    0    1    0    1    1    0    0    0    1    0    1    1    1    1    1    1    1    1    1    1    1
1    0    1    0    1    0    0    1    0    1    0    0    0    1    1    0    0    0    0     23
    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    0    0    1    1    1    1    1    1    1    1    1    1    1
1    1    1    1    1    1    1    1    1    1    0    1    1    1    0    0    1    0    1     41
    1    0    0    0    1    1    1    1    1    1    1    1    0    0    1    1    0    1    1    1    1    1    1    1    1    1    1    1
1    1    0    1    1    0    1    1    1    1    1    1    0    1    0    1    1    0    0     35
    0    0    0    0    0    0    1    1    1    0    0    1    0    0    0    0    0    0    0    1    1    1    1    1    1    1    1    0
1    0    0    1    0    1    1    0    0    0    0    0    1    0    0    0    1    0    0     18
    0    0    0    1    1    0    0    0    1    0    1    0    0    0    0    1    0    1    1    1    1    1    1    1    0    1    0    0
0    1    1    0    0    0    0    1    1    0    0    0    0    0    0    0    1    0    0     18
    1    1    1    1    0    0    0    1    1    1    1    1    1    0    0    1    1    1    1    1    1    1    1    1    1    0    1    0
1    0    1    1    1    0    1    0    1    0    0    0    0    0    1    0    1    0    0     29
```

```
    0    1    1    0    0    0    0    1    1    0    1    0    0    1    0    0    0    0    1    1    1    1    1    1    1    1    1    0
0    0    1    0    0    1    1    0    0    1    0    0    0    0    0    0    0    0    0     19
    0    0    0    0    1    0    1    0    0    0    1    0    0    0    0    0    0    0    1    0    1    1    1    1    1    0    0    0
1    1    0    0    1    0    0    0    0    0    0    0    0    1    0    0    0    0    1     14
    0    1    0    1    0    0    0    1    0    0    0    1    0    0    1    0    0    0    0    0    1    1    1    1    1    0    1    1
1    1    0    1    1    0    1    0    0    0    0    0    0    0    0    0    0    0    0     17
    0    0    0    0    0    0    0    0    1    0    0    0    0    0    0    0    0    0    0    0    1    1    0    1    0    1    0    0
1    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    1    0    0       7
    0    0    0    1    0    0    0    0    0    0    1    1    0    0    0    0    0    1    0    1    1    1    1    1    1    1    0    1
1    1    1    0    1    0    1    0    1    1    0    1    1    1    1    0    1    0    0     24
    0    1    0    1    0    0    0    0    1    0    0    1    0    0    0    0    0    1    0    0    1    1    1    0    1    1    1    0
0    1    0    0    0    0    0    0    1    0    0    0    0    0    0    0    0    0    0     13
    0    0    0    1    0    1    0    0    0    0    1    0    0    0    0    0    0    0    0    0    1    1    1    1    1    1    1    1
1    0    1    0    0    0    0    1    0    0    1    0    1    0    0    0    1    0    1     18
    0    1    0    0    0    0    0    0    1    0    0    1    0    1    1    1    0    0    0    1    1    0    1    1    0    1    1    1
0    1    1    0    0    0    1    1    1    0    0    0    0    1    0    0    0    0    0     19
    0    1    0    0    1    0    0    0    0    0    1    1    1    1    1    1    0    0    0    1    1    1    1    1    1    1    0    1
1    1    0    1    1    1    1    0    1    0    0    1    0    1    0    0    0    1    0     26
    0    1    1    1    1    1    1    0    1    0    1    1    1    0    0    1    0    1    0    1    1    1    1    1    1    1    1    1
1    1    0    1    1    1    0    0    1    1    0    1    0    0    0    0    0    1    0     30
    0    1    0    0    1    0    0    0    1    1    1    1    1    1    0    1    0    1    0    1    1    1    0    1    1    1    0    1
0    0    1    0    1    0    1    1    0    0    0    1    0    0    0    0    1    0    0     23
    1    0    1    1    1    1    1    1    1    1    1    1    1    1    1    0    1    1    1    1    1    1    1    1    1    1    1    1
1    0    0    1    1    0    1    1    0    1    0    0    0    0    1    0    0    1    1     35
    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    0    1    0    1    1    1    1    1    1    1
1    1    0    1    1    1    1    1    0    1    1    0    1    1    0    0    1    0    0     38
    0    0    0    0    1    0    0    0    1    0    1    1    1    0    1    1    0    0    0    0    1    1    0    1    1    1    1    1
0    0    0    1    1    0    0    1    1    1    0    1    0    1    0    0    1    1    0     23
    0    1    0    1    1    0    1    1    1    1    1    0    1    1    0    1    0    0    0    1    1    1    1    0    1    1    1    1
1    0    1    1    1    1    0    1    0    1    0    0    0    1    1    1    0    0    0     29

   12   21   14   20   20   16   17   21   27    16   32    31 24   12    17 23     5   20   22   28   40   39   37   36   37   35   25   29    29 23    22   24   22   20   22
19   22   20     7   17   12    15    10    6    19    9     6
```

Footnotes

[1] We thank A. Fink, M. Fritz, M. Kraxner, M. Müller, U. Sampt, C. Schermann, M. Singer, B. Suschnig, and W. Wagner for their assistance in collecting the data.

[2] We would like to express our gratitude to Aljoscha Neubauer who generously provided his data for this reanalysis.

[3] The complete answer patterns of all participants are given in Appendix B and C.

[4] The symmetrical distance *d* between two sets A and B is defined as follows:

$d(\mathrm{A},\mathrm{B}) = |\mathrm{A}\ \Delta\ \mathrm{B}|$, where $\mathrm{A}\ \Delta\ \mathrm{B} = (\mathrm{A}\backslash\mathrm{B}) \cup (\mathrm{B}\backslash\mathrm{A})$.

[5] The simulations were conducted with the program RandDist (Musch, 1995). Distances were computed with the program Di (Held, 1994). The program BConstra (Hockemeyer, 1994) was used to construct the knowledge spaces.

[6] Symmetrical distances of random answer patterns.

[7] Threshold for the percentage of items that have to be solved in order to conclude that a participant is capable of solving the items of this problem type.

[8] Results with the D(NxTxM) model are identical for the 75% and 90% threshold condition because a problem type is always represented by three items in this model, and already with a threshold of 75%, all three items of the given problem type have to be solved to conclude that a participant is capable of solving the items of this problem type.

[9] The simple deterministic procedure does not take into account possible answer failures, i.e., lucky guesses and careless errors. Stochastic procedures to overcome this problem have been developed by Falmagne and Doignon (1988) and Unnewehr (1992; Lukas & Unnewehr, 1993).

Table 1

Frequency of occurence of different numbers of rules (N), types of rules (T) and material attributes (M) in the 46 items of the APM

		Frequency of occurrence			
		1x	2x	3x	4x
Number of Rules (N)					
1		14			
2		17			
3		10			
4		5			
Type of Rule (T)					
O	Constant in a row	6	11	1	1
O	Pairwise progression	7	1	1	0
O	Distribution of two values	5	3	3	1
O	Distribution of three values	3	2	0	0
O	Figure Addition	11	1	0	0
O	Figure Subtraction	2	0	0	0
X	Exclusive-or	3	1	0	0
Material Attribute (M)					
L	Geometrical Figures	27	2	4	0
L	Patterns	9	5	3	1
L	Number	6	3	0	0
H	Spatial order	9	1	0	0

Table 2

Frequency of occurence of APM items with different demand components in the 46 items of Set I and II.

	Low Material Difficulty		High Material Difficulty	
Number of Rules	no X-or involved	X-or included	no X-or involved	X-or included
1	(1,O,L) *n* = 12 I: 3,4,10,12. II: 3,4,7,9,10, 11,12,16.	(1,X,L) *n* = 2 I: - II:22,23.	(1,O,H) *n* = 0 I: - II: -	(1,X,H) *n* = 0 I: - II: -
2	(2,O,L) *n* = 13 I:5,6,7,8,11 II:1,2,5,6,8, 15,19,20.	(2,X,L) *n* = 1 I: - II:33.	(2,O,H) *n* = 3 I: - II:14,25,26.	(2,X,H) *n* = 0 I: - II: -
3	(3,O,L) *n* = 5 I: - II:17,21,30,31,35.	(3,X,L) *n* = 1 I: - II:36.	(3,O,H) *n* = 4 I:9. II:13,18,34.	(3,X,H) *n* = 0 I: - II: -
4	(4,O,L) *n* = 2 I: - II:24,28.	(4,X,L) *n* = 0 I: - II: -	(4,O,H) *n* = 3 I: - II:27,29,32.	(4,X,H) *n* = 0 I: - II: -

Table 3

A list of the models that result from a dominance rule or a lexicographic rule between different problem components

Dominance rule based models	Lexicographic rule based models
D (N x T x M)	L (N > T > M)
	L (N > M > T)
D (N x T)	L (T > M > N)
D (N x M)	L (T > N > M)
D (T x M)	L (M > N > T)
	L (M > T > N)
D (N) = L (N)	
D (T) = L (T)	L (N > T)
D (M) = L (M)	L (T > N)
	L (N > M)
	L (M > N)
	L (T > M)
	L (M > T)

Table 4

Properties and goodness of fit of five models that best describe the aggregate data of Kratzmeier (1976) as measured on the level of individual answer patterns with two empirical data sets obtained with the original APM items, and a third data set obtained with newly constructed APM items. A simulation of 1000 random answer patterns was also conducted for each model.

							Distances										
Model	number of states	number of nonstates	data set		number of different states observed	number of different nonstates observed	0	1	2	3	4	5	6	7	8	mean symmetrical distance	distance agreement coefficient (DA)
D (N)	5	11	I	Random[6]			13	22.9	5.1	0	0	0	0	0	0	.81	-
				50%[7]	5	1	39	2	0	0	0	0	0	0	0	.05	.06
				75 %	5	2	39	2	0	0	0	0	0	0	0	.05	.06
				90 %	5	3	38	3	0	0	0	0	0	0	0	.07	.09
	5	11	II	Random			13.6	24.8	5.6	0	0	0	0	0	0	.81	-
				50 %	5	3	37	7	0	0	0	0	0	0	0	.16	.20
				75 %	5	2	40	4	0	0	0	0	0	0	0	.09	.11
				90 %	3	4	39	5	0	0	0	0	0	0	0	.11	.14
	5	11	III	Random			12.5	22.5	5	0	0	0	0	0	0	.81	-
				50 %	5	5	34	6	0	0	0	0	0	0	0	.15	.18
				75 %	5	3	34	6	0	0	0	0	0	0	0	.15	.18
				90 %	3	3	36	4	0	0	0	0	0	0	0	.10	.12

D (N x T)	14	114	I	Random			4.5	14.7	16.3	5.5	0	0	0	0	0	1.41	-
				50 %	9	7	33	8	0	0	0	0	0	0	0	.20	.14
				75 %	11	7	29	12	0	0	0	0	0	0	0	.29	.21
				90 %	8	13	23	14	4	0	0	0	0	0	0	.54	.38
	14	114	II	Random			4.9	15.8	17.4	6.0	0	0	0	0	0	1.41	-
				50 %	11	9	34	8	2	0	0	0	0	0	0	.27	.19
				75 %	9	9	31	11	2	0	0	0	0	0	0	.34	.24
				90 %	6	9	24	15	4	1	0	0	0	0	0	.59	.42
	15	241	III	Random			2.4	10.1	15.7	10.1	1.7	0	0	0	0	1.97	-
				50 %	10	12	25	11	3	1	0	0	0	0	0	.50	.25
				75 %	3	16	19	16	5	0	0	0	0	0	0	.65	.33
				90 %	2	6	32	5	2	1	0	0	0	0	0	.30	.15

(Table continued on next page)

D (T x M)	5	3	I	Random			25.6	15.4	0	0	0	0	0	0	0	.38	-
				50 %	5	0	41	0	0	0	0	0	0	0	0	.00	.00
				75 %	5	0	41	0	0	0	0	0	0	0	0	.00	.00
				90 %	5	3	37	4	0	0	0	0	0	0	0	.10	.26
	5	3	II	Random			27.6	16.4	0	0	0	0	0	0	0	.38	-
				50 %	5	0	44	0	0	0	0	0	0	0	0	.00	.00
				75 %	5	1	42	2	0	0	0	0	0	0	0	.05	.12
				90 %	3	1	42	2	0	0	0	0	0	0	0	.05	.12
	6	10	III	Random			15.1	20.0	5.0	0	0	0	0	0	0	.75	-
				50 %	6	3	35	5	0	0	0	0	0	0	0	.13	.17
				75 %	5	0	40	0	0	0	0	0	0	0	0	.00	.00
				90 %	4	1	39	1	0	0	0	0	0	0	0	.03	.03

D (NxTxM)	37	987	I	Random			1.5	7.2	14.0	12.7	5.2	0.4	0	0	0	2.35	-
				50 %	10	15	20	20	1	0	0	0	0	0	0	.54	.23
				75 %	10	23	13	19	7	2	0	0	0	0	0	.95	.40
				90 %	8	22	14	16	10	1	0	0	0	0	0	.95	.40
	37	987	II	Random			1.7	7.6	15.0	13.8	5.5	0.4	0	0	0	2.34	-
				50 %	6	19	16	24	2	2	0	0	0	0	0	.77	.33
				75 %	8	18	17	21	3	3	0	0	0	0	0	.82	.35
				90 %	5	15	19	19	5	1	0	0	0	0	0	.73	.31
	105	65431	III	Random			0.1	0.6	2.6	6.3	10.0	10.6	6.9	2.5	0.3	4.49	-
				50 %	5	33	5	1	14	13	5	2	0	0	0	2.45	.55
				75 %	6	33	11	14	6	5	3	1	0	0	0	1.45	.32
				90 %[1]	6	33	11	14	6	5	3	1	0	0	0	1.45	.32
L (T > M)	4	4	I	Random			20.5	20.5	0	0	0	0				.50	-
				50 %	3	1	40	1	0	0	0	0	0	0	0	.02	.05
				75 %	4	1	36	5	0	0	0	0	0	0	0	.12	.24
				90 %	4	4	36	5	0	0	0	0	0	0	0	.12	.24
	4	4	II	Random			22.2	21.8	0	0	0	0	0	0	0	.50	-
				50 %	4	2	40	4	0	0	0	0	0	0	0	.09	.18
				75 %	4	2	29	15	0	0	0	0	0	0	0	.34	.68
				90 %	2	1	40	4	0	0	0	0	0	0	0	.09	.18
	5	11	III	Random			12.3	22.7	5.0	0	0	0	0	0	0	.82	-
				50 %	5	4	31	9	0	0	0	0	0	0	0	.23	.28
				75 %	4	1	37	3	0	0	0	0	0	0	0	.08	.09
				90 %	3	2	37	3	0	0	0	0	0	0	0	.08	.09

[1] Results with the D(NxTxM) model are identical for the 75% and 90% threshold condition because a problem type is always represented by three items in this model, and already with a threshold of 75%, all three items of the given problem type have to be solved to conclude that a participant is capable of solving the items of this problem type.

Figure Captions

Figure 1. A problem illustrating the format of APM items. This problem is isomorphic to the actual item 7 in Set I. (The correct answer is the second cross in the lower line of distractors).

Figure 2. Hasse diagrams for two examples of orders on an item set.

Figure 3. The Hasse diagram for the model D(NxM).

Figure 4. The Hasse diagram for the model L(N>M).

Figure 5. The Hasse diagram of the model D(N x T). *n* denotes the number of items of the depicted type; the bold numbers represent solution percentages observed by Kratzmeier (1976) in a large reference sample of 1015 Germans of both sexes.

Figure 6. The Hasse diagram of model D(N). *n* denotes the number of items of the depicted type; the bold numbers represent solution percentages observed by Kratzmeier (1976) in a large reference sample of 1015 Germans of both sexes.

Figure 7. The Hasse diagram of model D(TxM). *n* denotes the number of items of the depicted type; the bold numbers represent solution percentages observed by Kratzmeier (1976) in a large reference sample of 1015 Germans of both sexes.

Figure 8. The Hasse diagram of the model L(T > M). *n* denotes the number of items of the depicted type; the bold numbers represent solution percentages observed by Kratzmeier (1976) in a large reference sample of 1015 Germans of both sexes.

Figure 9. The Hasse diagram of the least restrictive model D(NxTxM). *n* denotes the number of items of the depicted type; the bold numbers represent solution percentages observed by Kratzmeier (1976) in a large reference sample of 1015 Germans of both sexes. Problem types that are not represented by at least one APM item are omitted.

Figure 1

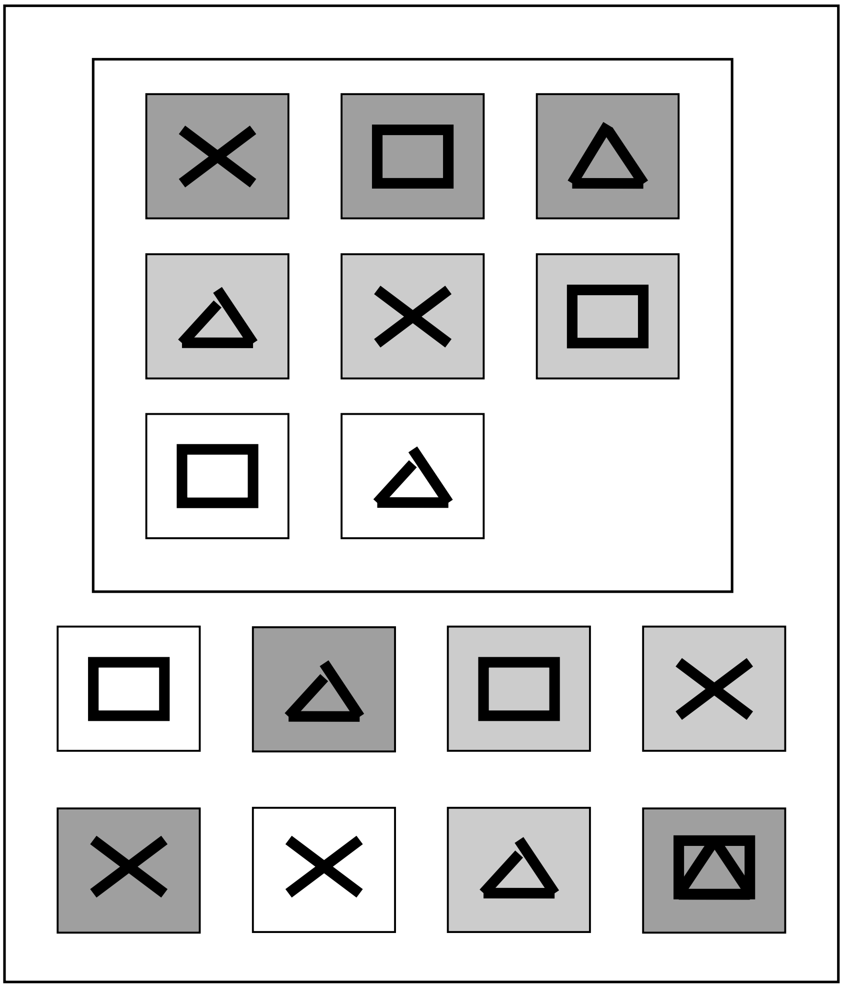


Figure 2

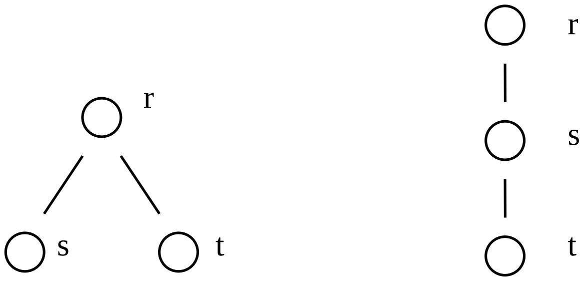


Figure 3

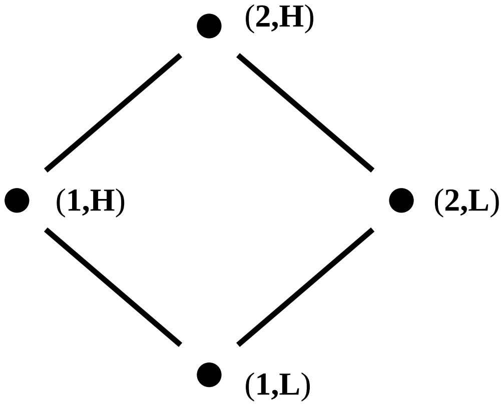


Figure 4

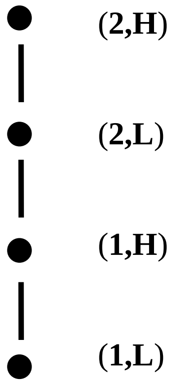


Figure 5

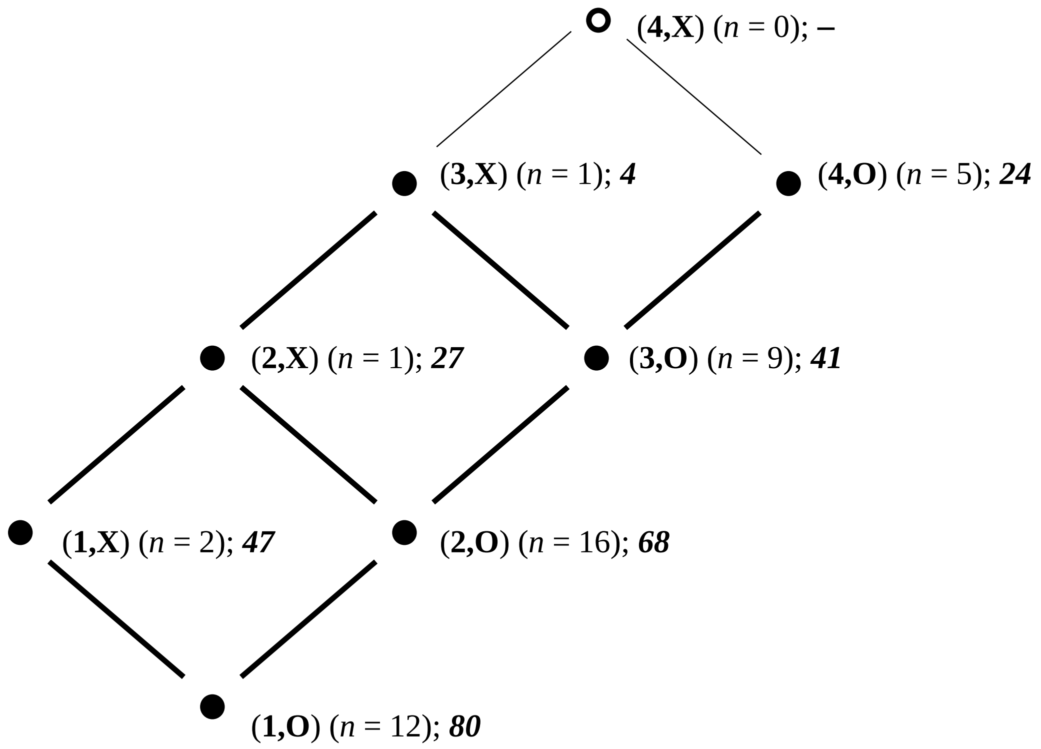

(**4,X**) (*n* = 0); **–**
(**3,X**) (*n* = 1); ***4***
(**4,O**) (*n* = 5); ***24***
(**2,X**) (*n* = 1); ***27***
(**3,O**) (*n* = 9); ***41***
(**1,X**) (*n* = 2); ***47***
(**2,O**) (*n* = 16); ***68***
(**1,O**) (*n* = 12); ***80***

Figure 6

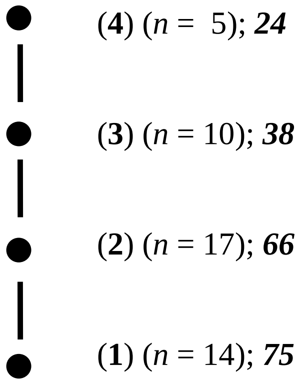

(**4**) (*n* = 5); ***24***
(**3**) (*n* = 10); ***38***
(**2**) (*n* = 17); ***66***
(**1**) (*n* = 14); ***75***

Figure 7

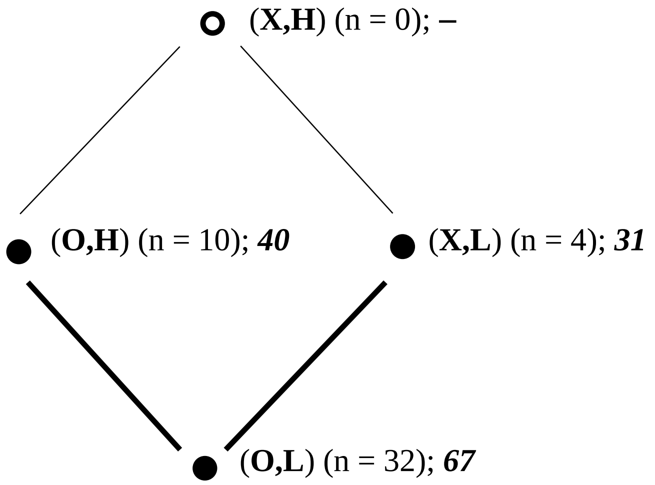

(**X,H**) (n = 0); **–**
(**O,H**) (n = 10); ***40***
(**X,L**) (n = 4); ***31***
(**O,L**) (n = 32); ***67***

Figure 8

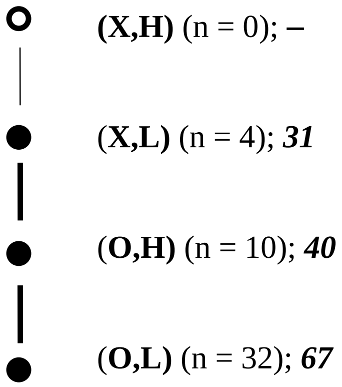
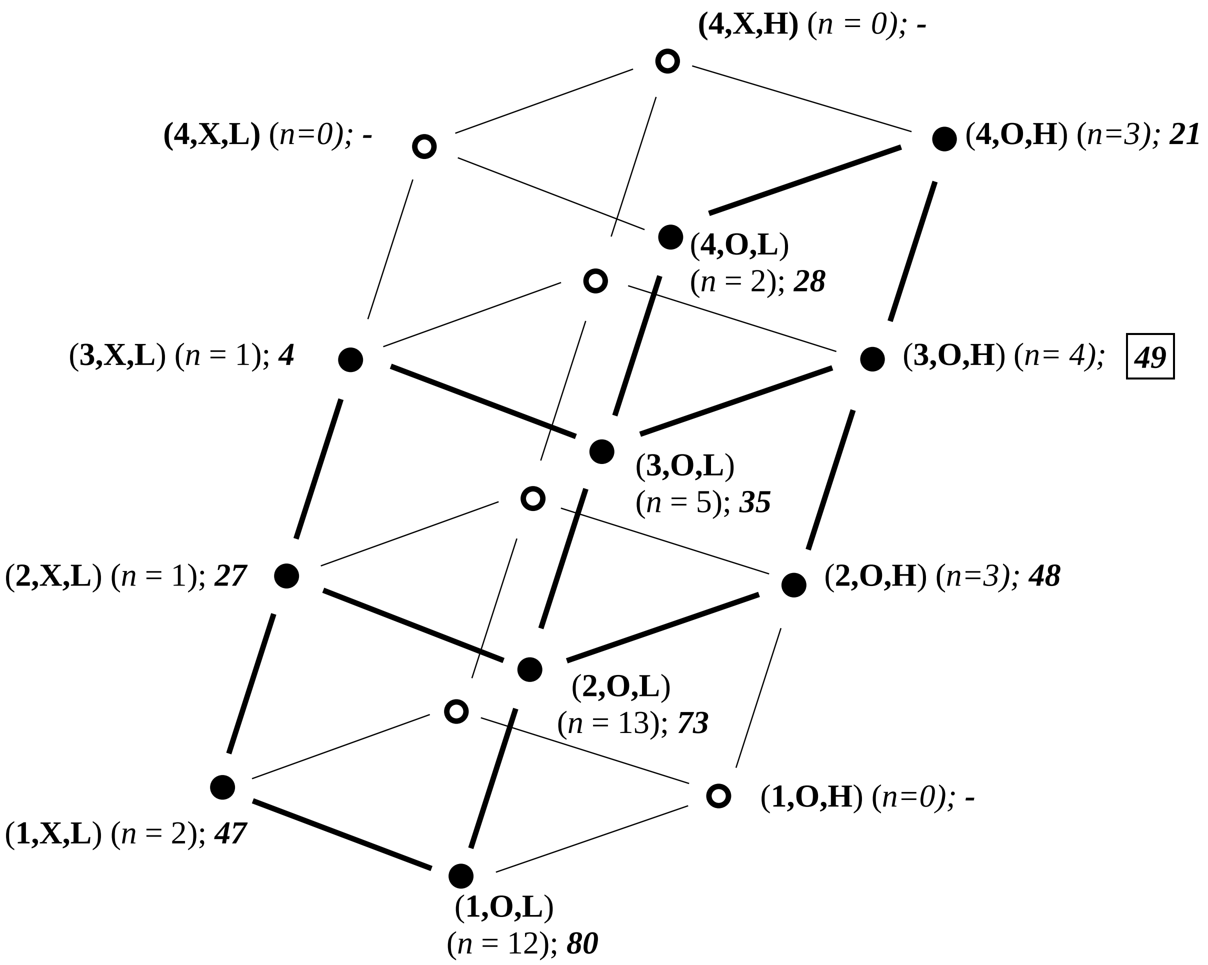


Figure 9


(4,X,H) (*n = 0); -*
(4,X,L) (*n=0); -*
(**4,O,H**) (*n=3); **21***
(**4,O,L**)
(*n* = 2); ***28***
(**3,X,L**) (*n* = 1); ***4***
(**3,O,H**) (*n= 4);* ***49***
(**3,O,L**)
(*n* = 5); ***35***
(**2,X,L**) (*n* = 1); ***27***
(**2,O,H**) (*n=3); **48***
(**2,O,L**)
(*n* = 13); ***73***
(**1,O,H**) (*n=0); -*
(**1,X,L**) (*n* = 2); ***47***
(**1,O,L**)
(*n* = 12); ***80***