CATSCALE: A stochastic multidimensional scaling methodology for the spatial analysis of sorting data and the study of stimulus categorization

Wayne S. DeSarbo

School of Business Administration, University of Michigan, Ann Arbor, MI, USA

.

Robert Libby

Johnson Graduate School of Management, Cornell University, Ithaca, NY, USA

Kamel Jedidi

Graduate School of Business, Columbia University, New York, NY, USA

Abstract: Sorting tasks have provided researchers with valuable alternatives to traditional paired-comparison similarity judgments. They are particularly well-suited to studies involving large stimulus sets where exhaustive paired-comparison judgments become infeasible, especially in psychological studies investigating stimulus categorization. This paper presents a new stochastic multidimensional scaling procedure called CATSCALE for the analysis of three-way sorting data (as typically collected in categorization studies). We briefly present a review of the role of sorting tasks, especially in categorization studies, as well as a description of several traditional modes of analysis. The CATSCALE model and maximum likelihood based estimation procedure are described. An application of CATSCALE is presented with respect to a behavioral accounting study examining auditors' categorization processes with respect to various types of errors found in typical financial statements.

Correspondence to: Professor W. DeSarbo, School for Business Administration, University of Michigan, Ann Arbor, MI, USA.

1. Introduction

Sorting tasks provide researchers with valuable alternatives to traditional paired-comparison similarity judgments. They are particularly well-suited to studies involving large stimulus sets where exhaustive paired comparison judgments are infeasible (Rao and Katz, 1971). Asking subjects to provide excessive paired comparisons may result in fatigue and alter the basis of their judgments (Johnson, Lehmann, and Horne, 1990). Having subjects sort the stimuli into piles of similar alternatives greatly simplifies the subjects' task.

Moreover, a sorting of stimuli based on perceived category membership is a natural reaction to large stimulus sets. Rosch (1975), and Mervis and Rosch (1981) define categorization as the equal treatment of items that can be discriminated on the basis of their attributes. Humans form categories to simplify their inherently complex world. In an attempt to map natural categorizations or objects, actions, or relationships, subjects are normally asked to sort the stimuli on the basis of some similarity criterion. The definition of the similarity criterion is normally decided upon by the subjects. The stimuli in such categorization studies have included mathematics problems (Schoenfeld and Herrmann, 1982), semantic relations (e.g., agent–activity, object–function, etc.; Chaffin and Herrmann, 1984), personality attributes (Andersen and Klatzky, 1987), etc. The purposes of these studies range widely from comparing the natural categorizations of novices and experts (or adults and children), to validating a priori categorizations based on some underlying theory.

The traditional approach to analyzing sorted stimuli is to initially calculate some similarity measure. For example, the relative frequency with which subjects sort any two stimuli together in the same subset may be treated as a measure of their similarity. This pooled or aggregate measure is then analyzed using existing two-way multidimensional scaling or clustering techniques. Yet, there are at least two important limitations to this approach. First, the input data may not be compatible with the technique. Some very unique aspects of the subjects' perceptions may be lost in the aggregation or pooling of the subjects' sorting data into a continuous similarity measure and its subsequent scaling or clustering. For example, subjects who form more natural categories will have greater impact on the solution than subjects who display less fine discrimination processes. Second, because these continuous measures result from an aggregation of responses across presumably homogeneous individuals, individual differences are hidden. This loss of individual differences occurs at two levels. Assuming subjects use the same aspects or attributes to judge stimuli, the weight placed on the aspects will likely vary from subject to subject. In judging cookies, for example, some subjects may give more weight to sweetness while others may give more weight to texture. Furthermore, the data may represent an averaging of two or more qualitatively different cognitive structures. For example, one consumer may distinguish cookies exclusively on sweetness while a second may distinguish the same cookies exclusively on texture. The structure revealed by a traditional multidimensional scaling or clustering analysis of aggregated relative

frequencies may bear little resemblance to either subject's actual cognitive structure. In the following section, we describe a methodology for the analysis of sorting task data via a new multidimensional scaling (MDS) procedure designed to overcome these potential problems.

2. The CATSCALE methodology

Many of the categorization studies performed in the social sciences utilize a common data collection procedure. Basically, the entire set of objects is placed on cards, one object per card. The subject is then asked to sort the cards into as many or few piles as he/she desires, congruent with the similarities or dissimilarities perceived among the stimuli. The subject is then asked to place the cards (stimuli) that they view as similar in the same pile. Later, individual level information is collected (e.g., experience/expertise, usage, attribute ratings, demographics, etc.) and related to the particular scheme used to form these 'piles'. Given such a data collection procedure, we are proposing a new methodology to examine the underlying perceptual categorization process. The data collected is 'subject conditional' in the sense that subjects may utilize differing numbers of piles/groups, as well as differing numbers and types of objects per pile/group. We wish to devise a new MDS based methodology which derives a common space for the objects from such data, as well as individual level parameters that can be directly interpreted in terms of individual differences in subjects' categorization processes.

A. The model

We ask subjects to subjectively group a designated set of stimuli into piles of similar objects on whatever basis they desire, where the number of piles is determined by the individual subject. Let

$i = 1, \ldots, I$ subjects;

$j, k = 1, \ldots, N$ stimuli or objects;

$t = 1, \ldots, T$ dimensions;

$\delta_{ijk} = \begin{cases} 1 & \text{if subject } i \text{ sorts objects } j \text{ and } k \text{ in the same pile,} \\ 0 & \text{otherwise;} \end{cases}$

$X_{jt} = $ the t-th coordinate of the j-th object;

$W_{it} = $ the importance or salience of the t-th dimension to subject i;

$\gamma_i = $ a threshold parameter for subject i.

We define a latent, unobservable squared distance function

$$D_{ijk} = d_{ijk} + e_{ijk},$$ \hfill (1)

where

$$d_{ijk} = \sum_{t=1}^{T} W_{it}(X_{jt} - X_{kt})^2,$$ (2)

and

e_{jk} = the error term, where it is assumed that $e_{ijk} \sim$ IID $N(0, \sigma_i^2)$.

Note, expression (2) is of the form of Carroll and Chang's (1970) INDSCAL model where $X = ((X_{jt}))$ provides the coordinates of the N objects in T dimensions, and $W = ((W_{it}))$ depicts individual differences via differential dimensional salience by subject.

Now, in order for two objects to be sorted into the *same* pile by subject i, we assume that the latent squared distance term defined in (1) is of a quantifiable magnitude *below* γ_i; i.e.,

$$P(\delta_{ijk} = 1) = P(D_{ijk} \leqslant \gamma_i)$$

$$= P(d_{ijk} + e_{ijk} \leqslant \gamma_i)$$

$$= P\left(e_{ijk} \leqslant \gamma_i - \sum_{t=1}^{T} W_{it}(X_{jt} - X_{kt})^2\right)$$

$$= \Phi\left(\frac{\gamma_i - \sum_{t=1}^{T} W_{it}(X_{jt} - X_{kt})^2}{\sigma_i}\right)$$

$$\underset{=}{\text{WLOG}} \; \Phi\left(\gamma_i - \sum_{t=1}^{T} W_{it}(X_{jt} - X_{kt})^2\right)$$

$$= \Phi(\cdot) = P_{ijk}.$$ (3)

The variance term, σ_i, can be arbitrarily set equal to 1 since its value can be adsorbed in the numerator by the γ_i and W_{it} terms. Similarly,

$$P(\delta_{ijk} = 0) = P(D_{ijk} > \gamma_i) = 1 - P_{ijk}$$

$$= 1 - \Phi(\cdot).$$ (4)

Assuming independence over all i, j, and k subscripts, we can form the likelihood function

$$L = \prod_{i=1}^{I} \prod_{j \neq k}^{N} \Phi(\cdot)^{\delta_{ijk}}(1 - \Phi(\cdot))^{1 - \delta_{ijk}},$$ (5)

or the log-likelihood function

$$\log L = \sum_{i=1}^{I} \sum_{j \neq k}^{N} \left[\delta_{ijk} \log(\Phi(\cdot)) + (1 - \delta_{ijk}) \log(1 - \phi(\cdot))\right].$$ (6)
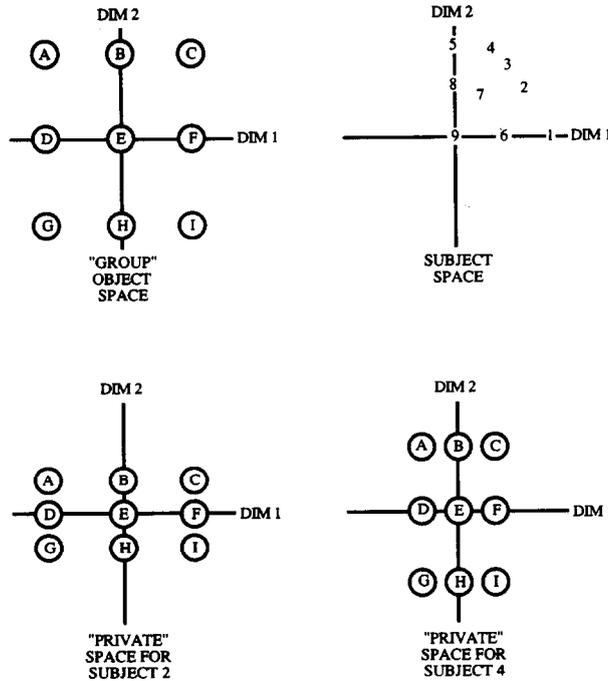
Fig. 1. Hypothetical example illustrating the INDSCAL/CATSCALE model (modified from Carroll and Wish, 1974, p. 62).

Thus, given $\Delta = ((\delta_{ijk}))$ and a specified value of T, the CATSCALE methodology estimates $X = ((X_{jt}))$, $W = ((W_{it}))$, and $\gamma = ((\gamma_i))$ by maximizing the log-likelihood function derived in expression (6). The Appendix describes the alternating conditional maximum likelihood (ACML) algorithm utilized for parameter estimation. Upon convergence of this estimation procedure, we obtain a T-dimensional spatial representation of the sorted objects in a common multidimensional scaling space (X), a T-dimensional representation of subjects in a separate multidimensional space that reflects individual differences in terms of differential dimension salience (W), and a set of threshold values which reflect the precision of discrimination between subjects (γ). Note, somewhat similar threshold model specifications have been presented by DeSarbo et al. (1987) for asymmetric binary proximity data, and by Jedidi and DeSarbo (1990) for pick any/N choice data. These formulations, however, utilize an unfolding model specification reflecting the fact that the row column objects in the binary data dealt with are different, and two sets of points need to be estimated: one set for the row objects and one set for the column objects.

Figure 1 presents a small synthetic illustration of this INDSCAL-like multidimensional scaling model. The two-dimensional 'group' object space in the upper left panel of this Figure depicts nine hypothetical objects, labeled A through I, in a lattice configuration, based on the X matrix of object coordinates. In the upper right panel, the W matrix of weights is plotted as a set of coordinates for

nine subjects denoted by the numbers 1–9. As long as these weights are positive (see the Appendix), they can be regarded graphically as stretching factors applied to the dimensions of the group object space. Thus, these differential weights produce for each subject a 'private' objective space by rescaling (differentially stretching and shrinking by a factor given by $\sqrt{W_{it}}$) the dimensions of the group object space. Thus, for example, the private spaces for subjects 2 and 4 can be recreated as in the lower section of Figure 1. As shown, subject 2 has greater weight for the first dimension than for the second dimension, while the situation is reversed for subject 4. Finally, the γ_i threshold coefficients (not shown in Figure 1) are also estimated denoting the precision of discrimination among the subjects.

B. Model degrees of freedom

The model degrees of freedom (df) are defined as the number of independent parameters to be estimated. The number of indeterminacies of the model is to be subtracted from the total number of parameters to be estimated to obtain the model degrees of freedom. Here, the CATSCALE model degrees of freedom are

$$df = T(I + N) - 2T + I. \tag{7}$$

There are T indeterminacies due to the fact that the origin of the coordinates of the stimuli (X) is indeterminant (one can add or subtract some constant c_t to each dimension of X and still not affect the latent distance values). Another T indeterminacies are due to the fact that one can multiply each column of X by any arbitrary positive constant r_t, and divide W by the square root of r_t, and not affect the latent distance function value.

C. Program options

CATSCALE can accommodate both external and internal analyses. In an external analysis, the procedure solves for the relevant set of non-fixed parameters. For example, if the user supplies X (which might be derived from another MDS analysis), then one would solve for W and γ. In an internal analysis, one solves for all of the designated parameters of the model (i.e., W, X, and γ).

As indicated previously, CATSCALE provides reparameterization options (DeSarbo and Rao, 1984, 1986; DeSarbo et al. 1982) whereby users can reparameterize or constrain the coordinates of the stimuli as linear functions of designated attributes or features. For example, if data are available on attributes for the stimuli, then one can reparameterize the coordinate, X_{jt}, as

$$X_{jt} = \sum_{m=1}^{M} H_{jm}\alpha_{mt}, \tag{8}$$

where H_{jm} is the value of attribute m for stimulus j, and α_{mt} measures the contribution of attribute m to dimension t. As a result of this reparameteriza-

tion, one obtains an MDS space where locations of the stimuli are expressed as linear functions of their respective attributes. The α_{mt} coefficients can also aid in interpreting the resulting dimensions, as well as delineating the contribution of these attributes on each of the derived dimensions.

Finally, one can estimate a common threshold parameter, $\gamma = ((\gamma))$, or threshold parameters that vary by subject (γ_i). Confirmatory analyses are also possible such as restricting W to be equal to the unit matrix of ones to examine the hypothesis of no individual differences.

3. Application: Uncovering the category structure of auditors' knowledge

Much of the recent research on the determinants of skilled performance in problem solving and decision making has focused on the influence of the structure of knowledge on performance. The ability to recognize patterns in data, speed of performance, efficiency of information search, recall and recognition performance, and other aspects of problem solving are affected by structural differences (Glaser and Chi, 1988). The importance of the development of category structures to differentiate and organize knowledge of concepts is well recognized in cognitive and social psychology (e.g., Smith and Medin, 1981; Andersen and Klatzky, 1987) as well as various applied areas including marketing (Alba and Hutchinson, 1987), medicine (Bordage and Zacks, 1984), aviation (Schvaneveldt et al., 1985), accounting (Frederick, 1986), etc.

Studies which uncover the manner in which natural categories are structured are necessary to understand a variety of memory retrieval phenomena and the development of expertise in problem solving. Changes in category structures will, for example, alter the manner in which consumers react to advertising, physicians relate diagnostic signs, and auditors search for explanations for audit findings. From a practical perspective, such findings can be indicate the most effective approach to convey advertising information, guide the teaching of clinical diagnosis, and aid the development of expert systems which lead auditors to more likely explanations for their audit findings. We discuss a recent study investigating this categorization process for a small group of professional auditors.

A. Study description

1. Study context
Auditing involves the examination by an independent certified public accountant of the fairness with which of a company's financial statements represents its financial position, results of operations, and cash flows. The process involves examining the accounting system and independently verifying the accuracy of the records it produces to insure that the resulting statements are free from material error. This examination is normally conducted on a test or sample basis.

The auditor's basic decision problem is often conceptualized as involving the diagnosis of financial statement errors. The knowledge necessary to perform this task involves the generation of potential diagnostic hypotheses which relate to possible financial statement errors and the different types of audit evidence which are indicative of those errors. Following the medical diagnosis literature, we assume that the possible error hypotheses are a principal organizing element of that knowledge. The current study examines the underlying structure of this financial statement error knowledge for experienced auditors. Attempts to date to develop models of the category structure have been incomplete and have not been subjected to rigorous empirical testing (e.g., Coakley and Loebbecke, 1985).

As Einhorn (1976) and Libby (1985) noted, it is through this network of financial statement error knowledge that search for additional evidence and the evaluation of the implications of evidence for further search and possible financial statement adjustments will be guided. Understanding the manner in which experienced auditors naturally categorize financial statement errors could direct the organization of evidence in documents to aid in the efficient evaluation of categories of potential errors. It also can provide a basis for training or the development of computer decision aids which help the inexperienced auditor learn to perform tasks more like their experienced counterparts. Attempts to gather archival statistical data that can be used by the individual auditor in the diagnostic process (e.g., archival frequency data from prior audits) can also be more effective (if the data is organized to fit the natural categorization of errors.

2. Subjects

The subjects were 20 staff auditors employed by one 'Big 8' CPA firm. Each auditor had completed between 6 and 18 months of practical experience. The experimental task was completed during a required one-week professional development seminar for all staff with that level of experience. Their cooperation was obtained by agreement with the executive office of the participating firm.

3. Stimuli

The stimuli used in the card sorting task were 35 financial statement audit differences or errors typically discovered in the audit of medium-sized manufacturing companies. The errors were selected in a two stage process. First, an a priori classification scheme was constructed based on standard textbooks (e.g., Arens and Loebbecke, 1988), the firm audit manual, and an extensive study of the occurrence of financial statement errors in medium sized manufacturing companies (Coakley and Loebbecke, 1985). The classification scheme with respective categories is presented in Table 1. The first aspect, transactions cycle, refers to the accounting subsystem which records transactions relating to a major business activity such as sales and cash receipts, purchases and cash disbursements, indebtedness or borrowing, etc. The second aspect, the audit

Table 1
Financial statement error aspects and categories

Transaction Cycle
1. Sales and cash receipts
2. Inventory
3. Purchases and payables
4. Payroll
5. Indebtedness
6. Investments

Audit Objective Violated
1. Proper cutoff (transactions recorded in proper period)
2. Only valid transactions are recorded
3. Proper valuation
4. Proper classification
5. Invalid transactions are not recorded

---

objective violated, refers to the attributes of financial statement accuracy such as recording of transactions in the proper period, proper valuation of transactions, etc. which apply to each transaction within every cycle.

Crossing these two aspects produces some 30 joint errors. The final set of errors were selected on the basis of firm managers' responses to free and error cued recall tasks. This resulted in a list of 35 errors. They are listed with their classification on the two a priori aspects in Table 2. Each of the 35 errors were printed separately on 2 × 6 cards.

4. Procedure

Each subject was presented with a large manila envelope containing one page of instructions, the 35 cards presented in random order, and several business envelopes. They were told that the cards contained a financial statement audit discrepancy typically discovered at medium-sized manufacturing clients. Their task was to sort the cards into two or more piles grouped in a manner such that each pile would consist of audit differences that 'go together'. Once sorted into piles, they were to place each pile in a separate business envelope and write a short title on the outside of the envelope. Once completed, they inserted these labeled business envelopes into the large manila envelope, and returned the manila envelope to the experimenter. The sorting data for each subject were converted to a 35 × 35 binary symmetric similarity matrix. Each matrix element took on the value of 1 if the two cards were sorted into the same pile, and 0 otherwise.

B. Traditional analyses

As previously mentioned, most MDS studies involving sorting data typically aggregate or pool the binary data over subjects, and apply a two-way metric or non-metric MDS procedure to these aggregated similarity counts. In this spirit,

we applied Torgerson's (1958) metric multidimensional scaling method and the ALSCAL (Takane, Young, and de Leeuw, 1978) two-way non-metric multidimensional scaling procedure. Table 3 presents the respective variance-accounting-for (VAF) statistics for $T = 1, \ldots, 5$ dimensions for these two methods. It is not clear from either of the two columns in Table 3 what the appropriate dimensionality should be. The VAF statistics appear to trail-off linearly for the Torgerson method, whereas there appears to be a somewhat large jump in VAF for $T = 5$ dimensions with the two-way non-metric ALSCAL procedure. The resulting spaces were not very interpretable from either method.

Another general approach that was attempted was to apply three-way metric and non-metric multidimensional scaling methods to the three-way $(20 \times 35 \times 35)$ binary data array $\mathbf{A}$. The last two columns in Table 3 present the VAF statistics for INDSCAL (Carroll and Chang, 1970) and a three-way, non-metric (subject conditional) ALSCAL analysis. As shown in this table, the respective VAF statistics were quite poor. There appears to be no obvious dimensionality to select on the basis of this table. In addition, the interpretation of the resulting spaces was very difficult.

C. CATSCALE analysis

The CATSCALE methodology was applied to the three-way, two-mode binary data in $T = 1, \ldots, 5$ dimensions with $\sigma_{ijk} = 1$, a varying threshold by subject, and an internal analysis without a reparameterization option. Table 4 presents the various goodness-of-fit statistics by dimension. On the basis of incremental improvement in quality of fit, and subsequent interpretation, the $T = 2$ dimensional solution was selected as most parsimoniously representing the structure in the data. (The Appendix discusses the difficulties in using the likelihood ratio test and AIC heuristic in this setting.)

Figure 2 presents the X configuration of the 35 errors listed in Table 2. To assess the overall relationship between the a prior classification scheme and that uncovered by the CATSCALE methodology, the two hypothesized constructs in Table 1 were coded as dummy variables as described in Table 5, and regressed on the two dimensions of the CATSCALE solution. The results of this analysis are represented in Table 6. Dimension I and dimension II, respectively, are strongly related to the audit objective and transaction cycle aspects presented in Table 1. For the audit objective dimension I, the intercept and all of the dummy variables except A_4 (which represents the 'invalid transactions are not recorded' objective) are significant, and the adjusted R^2 is 0.84. All five of the transaction cycle dummy variables and the intercept are significant with the transaction cycle dimension II, and the adjusted R^2 is 0.97. This analysis suggests that the CATSCALE methodology seemingly uncovered the hypothesized multidimensional structure in the sorting data. (Note, we purposely did *not* report an analysis reparameterizing X with this design matrix since, as discussed below, we wished to examine any differences uncovered apart from this apriori classification scheme.)

Table 2
Auditor's list of 35 errors

Error number	Code*		Plot value	Error
1	1	1	A	Next period's sales and accounts receivable were recorded in the current period
2	1	2	B	Goods returned by customers were underbooked.
3	1	3	C	Bad debt expense and allowance were understated.
4	1	4	D	Other revenues were recorded as sales.
5	1	5	E	Sales to valid customers were recorded more than once.
6	2	1	F	Inventory received next period was recorded in the current period.
7	2	2	G	Relief of inventory was not recorded for goods shipped.
8	2	3	H	Inventory was not written down to net realizable value.
9	2	3	I	Overhead was overapplied to inventory.
10	2	4	J	Expense items were improperly charged to inventory.
11	2	5	K	Inventory was booked but not received.
12	3	1	L	Current period's purchases on account are recorded next period.
13	3	1	M	Payments on account made next period were booked this period.
14	3	2	N	Purchases on account were not recorded.
15	3	2	O	Accrued operating expenses and payable were underrecorded.
16	3	3	P	Invoice prices of purchases were understated.
17	3	4	Q	Operating expenses were capitalized as plant and equipment.
18	3	5	R	Payments on account were recorded but not made.
19	4	1	S	Accrued payroll at year end was not recorded.
20	4	2	T	Compensated absences were not accrued.
21	4	2	U	Pension expense was not booked.
22	4	3	V	Payroll withholdings were understated.
23	4	4	W	Administrative payroll was improperly allocated to production.
24	4	5	X	Payroll was recorded for nonexistent employees.
25	5	1	Y	Current period's accrued interest expense on long-term debt was booked next period.
26	5	2	Z	Bank debt was not recorded.
27	5	3	1	Interest expense and bond discount amortization were understated.
28	5	4	2	Current portion of long-term debt was improperly included as noncurrent.
29	5	4	3	Capital leases were treated as operating leases.
30	5	5	4	Debt retirement payments were booked but not made.
31	6	1	5	Accrued interest income at year end was overbooked.
32	6	2	6	Sales of investment securities were underbooked.
33	6	3	7	Marketable equity securities were not written down to lower of cost or market.
34	6	4	8	Long-term investments were included in the current portfolio.
35	6	5	9	Fictitious investment securities were booked.

* The first digit indicates the transaction cycle and the second indicates the audit objective violated tied to Table 1.

Table 3
Results from traditional analyses (VAF)

T	2-way Torgerson MDS	2-way non-metric ALSCAL	3-way INDSCAL	3-way non-metric ALSCAL
1	0.124	0.021	0.018	0.022
2	0.239	0.049	0.036	0.091
3	0.338	0.076	0.053	0.125
4	0.420	0.111	0.065	0.144
5	0.498	0.288	0.078	0.161

Inspection of the stimulus plots in Figure 2 and the simple correlations presented in Table 4 provides further insights into the differences between the a priori classification and the auditors' natural categories. Concerning the audit objective dimension I, none of the errors which are invalid transactions (No. 5) or classification errors (No. 4) overlap on dimension I with members of any other category as seen in Figure 2. The same holds true for all members of the proper cutoff category (No. 1), except for item S (Accrued payroll at year end

Table 4
CATSCALE goodness-of-fit statistics

T	df	$-\ln L$	Pbc	Match	Phi	AIC
1	73	5518.21	0.259	0.839	0.211	11182.42
2	126	4269.81	0.478	0.856	0.359	8791.62
3	179	3817.68	0.519	0.863	0.405	7993.36
4	232	3042.24	0.650	0.887	0.544	6548.48
5	285	2686.61	0.697	0.899	0.605	5943.22

Correlations of Table 1 categories and the CATSCALE dimensions

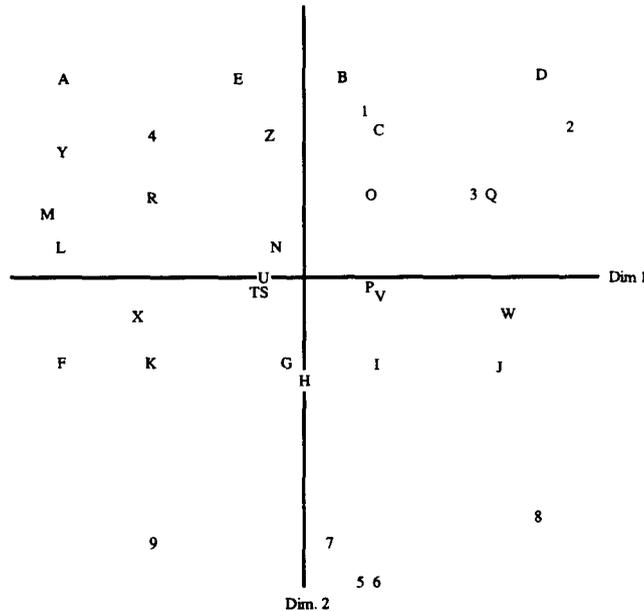| Category	Dimension	
	I	II
Transaction cycle		
1	0.043	0.530
2	-0.049	-0.316
3	-0.151	0.174
4	0.004	-0.077
5	0.055	0.414
6	0.118	-0.751
Audit objective		
1	-0.573	0.015
2	0.015	0.013
3	0.170	-0.093
4	0.747	0.061
5	-0.381	0.003

Fig. 2. CATSCALE two-dimensional solution for the auditor free-sort data.

was not recorded) which was categorized by subjects with the failure to record transactions. Although accrual errors result in timing errors on the financial statements, it appears that they are perceived instead as members of a category generated by a similar procedural error. The fit of these categories to the solution is supported by the R^2's in Table 6. Audit objective violations No. 2 and No. 3 (only valid transactions recorded and proper valuation), which had the lowest correlations with the solution, evidenced considerable overlap. The results suggest that errors in *amounts* recorded resulting in underrecording of

Table 5
Dummy variable coding for hypothesized aspects

Transaction cycle	T_1	T_2	T_3	T_4	T_5
1	0	0	0	0	0
2	1	0	0	0	0
3	0	1	0	0	0
4	0	0	1	0	0
5	0	0	0	1	0
6	0	0	0	0	1

Audit objective violated	A_1	A_2	A_3	A_4	
1	0	0	0	0	
2	1	0	0	0	
3	0	1	0	0	
4	0	0	1	0	
5	0	0	0	1	

Table 6
Regression analyses with X

	Dependent variable	
	X_1	X_2
Intercept	-1.01 *	3.39 **
T_1	-0.58	-3.58 **
T_2	-0.54	-1.72 **
T_3	-0.22	-2.66 **
T_4	-0.50	-0.70 **
T_5	0.39	-5.66 **
A_1	2.46 **	0.04
A_2	3.13 **	-0.12
A_3	5.55 **	0.02
A_4	0.61	0.05
RMSE	0.87	0.31
R^2	0.88	0.98
adj R^2	0.84	0.97
F-Ratio	19.96 **	131.28 **

 * $p \leqslant 0.05$.
 ** $p \leqslant 0.01$.

expense or overrecording of revenues are placed in one category, while errors involving *failure* to record a proper transaction are placed in another, indicating some discrepancy from the a prior classification structure. As suggested by the regression analysis, the relationship between the a priori transaction cycle aspect and dimension II is very strong. Only two of the 35 errors indicate category overlap (C and P), and this overlap is small. Thus, no changes in the a priori categorization structure on this dimension seem necessary. However, the differences between the a priori classification on audit objective and dimension I could provide the basis for a reorganization of teaching materials and audit workpaper organizations which may allow novice auditors to more quickly learn various audit tasks.

Finally, Table 7 presents the normalized weights and estimated threshold coefficients. As shown by a cursory inspection of these normalized weights, there appears to be significant individual differences in how these various auditors use these two dimensions for error categorization. For example, auditors 4, 5, 6, 9, 10, 13, 15, 16 and 18 possess high salience for the first dimension, while auditors 2, 7, 8, 11, 14 and 20 possess high salience for the second dimension. There is also evidence of several auditors that have large weights for both dimensions using more complex categorization processes. The threshold coefficients (γ_i) indicate that auditors 5, 10, and 14 utilize finer discrimination processes than the remaining 17 auditors given their larger γ_i coefficients. Unfortunately, because of the homogeneity of this subject group, we did not collect individual difference or background data on these 20 subjects to attempt

Table 7
CATSCALE estimated weights (normalized) and threshold coefficients

Subject i	W_{i1}	W_{i2}	γ_i
1	0.506	0.862	0.000
2	0.164	0.986	0.027
3	0.551	0.835	0.008
4	1.000	0.018	0.536
5	0.992	0.128	0.237
6	0.983	0.181	0.003
7	0.077	0.997	0.010
8	0.002	1.000	0.288
9	1.000	0.016	0.495
10	0.998	0.061	0.005
11	0.025	1.000	0.082
12	0.859	0.512	0.000
13	1.000	0.001	1.302
14	0.006	1.000	0.177
15	1.000	0.004	0.225
16	0.998	0.066	0.038
17	0.459	0.889	0.000
18	0.994	0.110	0.002
19	0.572	0.820	0.000
20	0.006	1.000	0.040

to explain these different categorizations. However, in future studies of groups with differing backgrounds, such analyses could be employed.

Appendix. The CATSCALE ACML procedure

Algorithm

In CATSCALE, the log of the likelihood function is maximized with respect to a specific set of parameters (i.e., the objects' coordinates $X = ((X_{jt}))$, the subjects' weights $W = ((W_{it}))$, and the threshold coefficients $\gamma = ((\gamma_i))$), holding the other sets constant. This conditional maximum likelihood estimation alternates across all the parameter sets until a convergence criterion is satisfied. Hence, we label the procedure ACML for Alternating Conditional Maximum Likelihood estimation. In each major iteration, CATSCALE utilizes a conjugate gradient method with automatic restarts (Powell, 1977) to estimate the particular parameter set. The major phases of this methodology are summarized below.

Phase 1. Input options

The user must specify the number of subjects (I), the number of objects (J), the number of dimensions (T), the type of analysis (internal or external), the

threshold option (varying by subject or constant), the type of starting solution (rational, given, or random), and the name of the input file containing the input data. This file must include the object coordinates if an external analysis is to be performed or if a given start is selected. In addition, it must also contain H if the reparameterization option is requested.

Phase 2. Starting estimates

Options exist for the user to specify a random start, a 'rational' start for X (a singular value decomposition, SVD, analysis on the double-centered, aggregated (over subjects) sorted data) or, a given start.

Phase 3. Estimate X, W and γ

We need to impose non-negativity constraints on the subject weights ($W_{it} \geq 0$) and threshold parameters ($\gamma_i \geq 0$). To accomplish this, we estimate the corresponding squared entities: W_{it}^2 and γ_i^2 as suggested by Gill, Murray, and Wright (1981, pp. 268–269).

Partial derivatives

Let

$$U_{ijk} = \sum_{t=1}^{T} W_{it}^2 (X_{jt} - X_{kt})^2, \tag{A.1}$$

$$\Phi(\cdot) = \Phi\left(\gamma_i^2 - U_{ijk}\right), \tag{A.2}$$

$$M_{ijk} = \left(\frac{1 - \delta_{ijk}}{1 - \Phi(\cdot)} - \frac{\delta_{ijk}}{\Phi(\cdot)}\right). \tag{A.3}$$

Then

$$\frac{\partial \ln L}{\partial W_{it}} = 2 \sum_{j > k}^{N} \sum W_{it}(X_{jt} - X_{kt})^2 \phi(\cdot) M_{ijk}; \tag{A.4}$$

$$\frac{\partial \ln L}{\partial X_{jt}} = 2 \sum_{i}^{I} \sum_{j \neq k}^{N} W_{it}^2 (X_{jt} - X_{kt}) \phi(\cdot) M_{ijk}; \tag{A.5}$$

$$\frac{\partial \ln L}{\partial \gamma_i} = -2 \sum_{j > k}^{N} \sum \gamma_i \phi(\cdot) M_{ijk}. \tag{A.6}$$

In the case of a reparameterized object space, we have the following partial derivatives with respect to the α_{mt} parameters:

$$\frac{\partial \ln L}{\partial \alpha_{mt}} = 2 \sum_{i}^{I} \sum_{j > k}^{N} \sum W_{it}^2 (X_{jt} - X_{kt})(H_{jm} - H_{km}) \phi(\cdot) M_{ijk}, \tag{A.7}$$

where $\phi(\cdot)$ represents the standard normal density function. At the start, the W_{ij}'s, the γ_i's, and the X_{jt}'s (in the case of an internal analysis with a random start option) are randomly generated from a uniform distribution. In the case of external analysis, the methodology solves for whatever parameter set(s) are not fixed/given. Holding W and γ constant, the algorithm uses the conjugate gradient method with automatic restarts (Powell, 1977) to estimate X (assuming an internal analysis). Then it estimates W, holding X and γ constant. Finally, γ is estimated while holding W and X fixed. The algorithm repeats these major phases until a convergence criterion is met.

With these partial derivatives specified, we can now briefly describe the conjugate gradient procedure with automatic restarts (Powell, 1977) for a given set of parameters.

(i) Start with initial parameter estimates $\Psi^{(1)}$, and set the iteration counter IC = 1. Ψ here denotes a vector, stacking the relevant set of parameters to be estimated (i.e., X, W, γ).

(ii) Set the first search direction $S^{(1)} = -\nabla(\ln L)^{(1)}$, where $\nabla(\ln L)^{(1)}$ denotes the gradient vector of the log-likelihood function evaluated at $\Psi^{(1)}$.

(iii) Find $\Psi^{(2)}$ using the relationship

$$\Psi^{(2)} = \Psi^{(1)} + \lambda^{(1)}S^{(1)}, \tag{A.8}$$

where $\lambda^{(1)}$ is the optimal step-size in the gradient direction $S^{(1)}$. A quadratic interpolation method is used for estimating the optimal step-size. Set IC = 2.

(iv) Calculate $\nabla(\ln L)^{(IC)}$ and set the (new) search direction

$$S^{(IC)} = -\nabla(\ln L)^{(IC)} + \beta S^{(IC-1)}, \tag{A.9}$$

if IC = 2 or if a restart is needed. In this algorithm, restarts are made every V iterations (V is the number of parameters to be estimated) or when the search direction is not 'sufficiently downhill'. If this is the case, then set $R =$ IC and go to step (vi). R holds the number of the iteration where a restart is made. Otherwise, set

$$S^{(IC)} = -\nabla(\ln L)^{(IC)} + \beta S^{(IC-1)} + \epsilon S^{(R)}, \tag{A.10}$$

where

$$\beta = \frac{\nabla(\ln L)^{(IC)'}\nabla(\ln L)^{(IC)}}{\nabla(\ln L)^{(IC-1)'}\nabla(\ln L)^{(IC-1)}}, \tag{A.11}$$

$$\epsilon = \frac{\left(\nabla(\ln L)^{(IC)} - \nabla(\ln L)^{(R)}\right)'\nabla(\ln L)^{(IC)}}{\left(\nabla(\ln L)^{(IC)} - \nabla(\ln L)^{(R)}\right)' S^{(R)}} \tag{A.12}$$

and

$S^{(R)} =$ the search direction when a restart is made.

(v) Check if $S^{(IC)}$ is 'sufficiently downhill' using the following condition:

$$-\Delta(\ln L)^{(IC)'} S^{(IC)} \geqslant \rho \| S^{(IC)} \| \| \nabla(\ln L)^{(IC)} \|, \tag{A.13}$$

where ρ is a positive constant. Powell (1977) suggests setting $\rho = 0.2$. If this condition is satisfied then go to step (vi), otherwise return to step (iv). This phase is performed only when $S^{(IC)}$ is computed via (A.10).

(vi) Compute the optimal step-size at iteration IC in the direction $S^{(IC)}$ and then compute:

$$\boldsymbol{\Psi}^{(IC+1)} = \boldsymbol{\Psi}^{(IC)} + \lambda^{(IC)}S^{(IC)}. \tag{A.14}$$

(vii) If $\boldsymbol{\Psi}^{IC+1)}$ is optimal, stop. Otherwise set $IC = IC + 1$ and return to step (iv). The optimality of $\boldsymbol{\Psi}^{(IC+1)}$ is determined using the following convergence criteria:

– The amount of improvement of the log-likelihood function between the last and before last iteration, i.e., $|\ln L^{(IC-1)} - \ln L^{(IC)}| \leqslant \text{TOL}$.
– The length of the gradient, i.e., $\|\nabla \ln L^{(IC)}\| \leqslant \text{TOL}$.
– The maximum number of iterations set by the user, i.e., $IC \geqslant IC^*$.

Phase 4. Normalization and output

Here, we normalize the dimensions (columns) of X to unit sums of squares and then redefine the matrix of situations' weights (W) to compensate for these transformations. Various goodness-of-fit measures (described below) are computed over the entire data, as well as by subject, to examine possible outliers in the data.

These goodness-of-fit measures, computed for a particular solution, are:

(1) The log-likelihood function;
(2) A deviance measure (Nelder and Weddenburn, 1972):

$$D = -2\left[\sum_{i=1}^{I} \sum_{j<k}^{N} \left[\delta_{ijk} \ln\left(\hat{P}_{ijk}\right) + \left(1 - \delta_{ijk}\right)\ln\left(1 - \hat{P}_{ijk}\right)\right]\right]$$

$$= -2 \ln L\left(\hat{P}_{ijk}\right). \tag{A.15}$$

This is basically (6) with P_{ijk} replaced by its estimated value, $\hat{P}_{ijk}$. The difference between two deviance measures corresponding to two nested models is asymptotically χ^2 distributed with degrees of freedom equal to the difference in the models' degrees of freedom. This difference can be used (theoretically) to test for dimensionality as well as for various model specifications because of the obvious nesting. However, one potential problem with using such a χ^2 test concerns the presence of incidental parameters in the likelihood function. In other words, the number of parameters to be estimated varies with the number of individuals, stimuli, and situations considered. In such a case, the MLE's are not consistent (Anderson, 1980), and thus the χ^2 test is not appropriate.

(3) The Akaike information criterion (Akaike, 1974):

$$\text{AIC} = -2 \ln L\left(\hat{P}_{ijk}\right) + 2 \text{ (number of independent parameters}$$

in the model).

This AIC measure can be utilized to test for dimensionality and model selection. According to Akaike (1974), the model with minimum AIC should be selected. However, as noted by Bozdogan (1987), the use of this AIC criterion tends to result in over-fitting certain models.

(4) A simple matching coefficient (Match) calculated between Δ and the predicted $\hat{\Delta}$, according to the threshold rule of the model being estimated.

(5) The point biserial correlation (Pbc) between $\hat{P} = ((\hat{P}_{ijk}))$ and Δ.

(6) The phi coefficient (Phi) calculated between Δ and $\hat{\Delta}$.

All these goodness-of-fit measures need to be inspected in determining the dimensionality of the space and testing for nested models, given the problems and difficulties in using the χ^2 test and the AIC.

References

Akaike, H., A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, AC-**19** (1974) 716–723.

Alba, J.W. and J.W. Hutchinson, Dimensions of consumer expertise, *Journal of Consumer Research*, **13** (1987) 411–454.

Anderson, E.B, *Discrete Statistical Models with Social Science Applications* (North-Holland, New York, 1980).

Andersen, S.M. and R.L. Klatzky, Traits and social stereotypes: Levels of categorization in person perception, *Journal of Personality and Social Psychology*, **53** (1987) 235–246.

Arens, A.A. and J.K. Loebbecke, *Auditing, an Integrated Approach*, (Prentice-Hall, Englewood Cliffs, NJ, 1988).

Bacon, M.K. and R.D. Ashmore, How mothers and fathers categorize descriptions of social behavior attributed to daughters and sons, *Social Cognition*, **3** (1985) 193–217.

Bordage, G. and R. Zacks, The structure of medical knowledge in the memories of medical students and general practitioners: Categories and prototypes, *Medical Education*, **18** (1984) 406–416.

Bozdogan, H., Model selection and Akaike's information criteria (*AIC*): The general theory and its analytical extensions, *Psychometrika*, **52** (1987) 345–370.

Carroll, J.D. and J.J. Chang, Analysis of individual differences in multidimensional scaling via an *N*-way generalization of Eckart–Young decompositions, *Psychometrika*, **35** (1970) 283–319.

Carroll, J.D. and M. Wish, Multidimensional perceptual models and measurement methods, in: E.C. Carterette and M.P. Friedman (Eds.), *Handbook of Perception: Vol. 2, Psychophysical Judgment and Measurement* (Academic Press, New York, 1974) pp. 391–447.

Chaffin, R. and D.J. Herrmann, The similarity and diversity of semantic relations, *Memory and Cognition*, **12** (1984) 134–141.

Coakley, J.R. and J.K. Loebbecke, The expectation of accounting errors in medium–sized manufacturing firms, Advances in Accounting, **2** (1985) 199–245.

DeSarbo, W.S. and V.R. Rao, GENFOLD2: A set of models and algorithms for the GENeral unFOLDing analysis of preference/dominance data, *Journal of Classification*, **1** (1984) 147–186.

DeSarbo, W.S. and V.R. Rao, A constrained unfolding methodology for product positioning, *Marketing Science*, **5** (1986) 1–19.

DeSarbo, W.S., J.D. Carroll, D.R. Lehmann and J. O'Shaughnessy, Three-way multivariate conjoint analysis, *Marketing Science*, **1** (1982) 323–350.

DeSarbo, W.S., D. Lehman, S. Gupta, M. Holbrook and W. Havlena, A three-way unfolding methodology for asymmetric binary proximity data, *Applied Psychological Measurement*, **11** (1987) 397–418.

Einhorn, H.J., A synthesis: Accounting and behavioral science, *Journal of Accounting Research*, **14** (1976) 196–206.

Frederick, D.M., Auditors' representation and retrieval of knowledge in internal control evaluation, Unpublished doctoral dissertation (University of Michigan, 1986).

Gill, P.E., W. Murray and M.H. Wright, *Practical Optimization* (Academic Press, Orlando, FL., 1981).

Glaser, R.H. and M.T.H. Chi, *Overview, in The Nature of Expertise* (Lawrence Erlbaum Associates, Hillsdale, N.J., 1988).

Jedidi, K. and W.S. DeSarbo, A stochastic multidimensional scaling procedure for the spatial representation of three-way, three-mode pick any/J data, *Psychometrika*, **56** (1994) 471–494.

Johnson, M.D., D.R. Lehmann and D.R. Horne, The effects of fatigue on judgments of interproduct similarity, *International Journal of Research in Marketing*, **7** (1990) 35–43.

Libby, R., Availability and the generation of hypotheses in analytical review, *Journal of Accounting Research*, **23** (1985) 648–667.

Mervis, C.B. and E. Rosch, Categorization of natural objects, *Annual Review of Psychology*, **32** (1981) 89–115.

Nelder, J.A., and R.W.M. Wedderburn, Generalized linear models, *Journal of the Royal Statistical Society Series A*, **135** (1972) 370–384.

Powell, M.J.D., Restart procedures for the conjugate gradient method, *Mathematical Programming*, **12** (1977) 241–254.

Rao, V.R. and R. Katz, Alternative multidimensional scaling methods for large stimulus sets, *Journal of Marketing Research*, **8** (1971) 488–494.

Rosch, E., Cognitive representation of semantic categories, *Journal of Experimental Psychology: General*, **104** (1975) 192–233.

Schoenfeld, A.H. and D.J. Herrmann, Problem perception and knowledge structure in expert and novice mathematical problem solvers, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **8** (1982) 484–494.

Schvaneveldt, R.W., F.T. Durso, T.E. Goldsmith, T.J. Breen, N.M. Cooke, R.G. Tucker and J.C. DeMaio, Measuring the structure of expertise, *International Journal of Man – Machine Studies*, **23** (1985) 699–728.

Smith, E.E. and D.L. Medin, *Categories and Concepts* (Harvard University Press, Cambridge, MA., 1981).

Takane, Y., F.W. Young and J. de Leeuw, Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features, *Psychometrika*, **42** (1977) 7–67.

Torgerson, W.S., *Theory and Methods of Scaling* (Wiley, New York, 1958).