The effects of fatigue on judgments of interproduct similarity

Michael D. Johnson *

School of Business Administration, University of Michigan, Ann Arbor, MI 48109, U.S.A.

Donald R. Lehmann

School of Business Administration, Columbia University, New York, NY 10027, U.S.A.

Daniel R. Horne

School of Business Administration, University of Michigan, Ann Arbor, MI 48109, U.S.A.

Similarity scaling often requires subjects to produce such a large number of judgments that fatigue may become a problem. Yet it remains unclear just how respondent fatigue affects similarity perceptions and resulting judgments. The present study uses a categorization perspective to examine the effects of fatigue on similarity judgments. The results suggest that subjects rely increasingly on category membership as they progress through a similarity judgment task.

1. Introduction

Marketing researchers apply a variety of similarity scaling techniques, including multidimensional scaling (Shepard, 1962; Kruskal, 1964), hierarchical clustering (Johnson, 1967), additive clustering (Shepard and Arabie, 1979), and additive tree scaling (Sattath and Tversky, 1977), to help them understand consumer perceptions of product or service alternatives (see, e.g., Arabie, Carroll, DeSarbo and Wind, 1981; Cooper, 1983; Green and Carmone, 1970; Green and Rao, 1972; Srivastava, Leone and Shocker, 1981). These applications typically require a relatively large number of products to construct meaningful product representations (Klahr, 1969).

Asking subjects to make too many product judgments may, however, affect the quality of those judgments (Sudman and Bradburn, 1982). Consider traditional applications in which respondents are asked to provide similarity ratings of product pairs. Because the number of ratings required for the analysis increases roughly as the square of the number of items, i.e., $\frac{1}{2}n(n-1)$, applications invʳ ʳ ing a large number of products or services may become prohibitive (Hauser and Koppelman, 1979).

Marketing researchers have recognized that there is a limit to the amount of "quality" information that can be collected from respondents. This has led to the development of procedures for reducing the number of required judgments, including the use of sorting tasks (Rao and Katz, 1971), cyclical designs (David, 1988), and the random deletion of pairs (Malhotra, Jain and Pinson, 1988). Using a cyclical design, for example, requires the presence of two conditions, where each stimulus element should appear an equal number of times in the data, and the reduced set of stimulus pairs should not be divisible into nonintersecting subsets (Kendall, 1955). In a five stimulus array, the rated stimulus pairs might include the first and second, second and third, third and fourth, fourth and fifth, and fifth and first stimulus elements.

At the same time, Malhotra, Jain and Pinson (1988) have demonstrated that configuration recovery using incomplete data becomes poorer as the proportion of data that is mis-

sing, or the number of stimuli involved, increases. Many applications will continue to call for a relatively large number of paired comparison ratings. Therefore, the potential exists for fatigue to seriously affect these ratings. Importantly, it remains unclear just how subject fatigue affects similarity judgments.

The present paper examines the effects of fatigue on similarity judgments. More specifically, we adopt a categorization perspective (Mervis and Rosch, 1981; Medin and Smith, 1984) to predict just how subjects' perceptions and judgments change as they progress through a similarity rating task. Categorization research suggests that one natural reaction to fatigue is for subjects to rely increasing on the categorical similarities and differences among products. After describing in more detail how repetition and fatigue may affect similarity judgments, we develop our hypotheses. An empirical study that tests the hypotheses is then reported.

2. Task repetition, adaptation, and fatigue

The traditional method for collecting similarity scaling data is to have respondents rate the overall similarity of each possible pair of products on a proximity scale. A number of studies have examined the reliability of these direct similarity judgments (Day, Deutscher and Ryans, 1976; Moore and Lehmann, 1982; Summers and MacKay, 1976; Weksel and Ware, 1967) and obtained mixed results. Yet, similarity scaling, in particular multidimensional scaling (MDS), remains popular (Cooper, 1983) and appears fairly robust to changes in a number of factors, including the metric employed (Green, 1975), the order of presentation of the stimuli (Jain and Pinson, 1976), and the embedding of stimuli in a stimulus domain (Malhotra, 1987). At the same time, marketing researchers recognize the often burdensome and boring nature of similarity judgment tasks and the fatigue that may result (Malhotra, 1987).

What is unclear is just how respondents adapt to fatigue. Dong (1983), for example, found that while missing and inconsistent responses tend to increase over time, fatigue does not appear to influence aggregate similarities or MDS solutions. However, Dong's study focused on aggregate data rather than changes in individual judgments.

At least two factors may affect an individual's judgments through the course of a similarity rating task: adaptation and fatigue. Initially, respondents become comfortable or adapt to thinking about the items. The subjects should become more comfortable using the scale and provide higher test–retest correlations. Further, the basis of the judgments themselves, whether common or distinctive features (Tversky, 1977) or product categorizations (Mervis and Rosch, 1981), should become well established. In other words, with early repetition respondents should "settle in" and provide more consistent judgments. We expect that this adaptation either occurs relatively quickly or is minimized, if not eliminated, by appropriate pre-task procedures (e.g., a warm-up task or prior acclimation to the stimulus set).

As respondents continue to progress through the task, fatigue may affect the judgments. One straightforward way subjects may adapt to fatigue is to adopt a single response value or restricted set of responses for rating the product pairs. Under this scenario one should observe systematic changes in the subjects' judgments. Most notably, the variance in the judgments should decrease through the course of the rating task. One would also expect a general decrease in the ability of similarity scaling techniques to capture or fit the judgments as the correlation between the subjects' perceptions and their ratings decreases.

A second possibility is that fatigue results in increasingly careless responses. Subjects may simply become sloppy as they tire. This carelessness should result in an increase in the

error variance of the subjects' responses as they progress through the task. Similar to using a restricted response set, one should observe a decrease in the ability of scaling techniques to fit the judgments. Under either scenario, the effects of fatigue should not be contingent on the nature of the stimuli.

Although both factors may come into play, a more task-oriented reaction to fatigue is the adoption of simpler product representations in order to finish the task more quickly or easily. Research on categorization (Mervis and Rosch, 1981; Medin and Smith, 1984) provides a useful framework for predicting just how this simplification occurs. This research suggests that there is a basic level of categorization at which categories are most differentiated from one another and where categories exhibit a singularly high degree of inclusiveness or similarity of their members (Murphy, 1982; Murphy and Smith, 1982; Rosch, 1975a, 1975b; Rosch et al., 1976). For example, the category "apples" is at a more basic level than either its immediate superordinate category (i.e., "fruit") or its subordinate categories (e.g., "Delicious apples" – Rosch et al., 1976).

Recent consumer research further suggests that traditional consumer product categories (e.g., soft drinks) represent a basic level of categorization for consumers (Johnson and Fornell, 1987). Consumers may, therefore, rely more exclusively on traditional product category membership in order to simplify the task and adapt to fatigue. An important implication is that fatigue should differentially affect the variance in consumers' similarity judgments depending on whether the same or different product categories are involved.

When judging products that are all members of the same traditional category, the variance in the subjects' judgments should decrease over time. Many of the more subtle within category differences that might be taken into account early in the task are likely ignored later on. In judging soft drinks, for example, consumers may begin by distinguishing alternatives on flavor, brand name, sweetness, and after-taste. As the task drags on, the overriding similarity of the category members should increase in importance relative to more subtle or minor within category differences.

When judging products from a more superordinate level, involving more than one traditional category or basic level distinction (e.g., beverages), judgment variance should increase as subjects progress through the task. Here fatigue should result in increased reliance on the distinctiveness of the categories themselves. Many of the more subtle or minor across category similarities that may be considered early in the task are likely ignored later on. The more salient category based similarities and differences should predominate as fatigue sets in. This should result in an increase in judgment variance over time. For example, early on consumers may recognize product similarities within as well as across categories, such as the similarity between soft drinks and fruit juices. As consumers progress through the task they may simply rely on the categorical difference between these stimuli.

Under this scenario, the effect of fatigue on judgment variance should be contingent on the brand versus category nature of the stimuli. Similarity variance should show more of a decrease for products from a single category and more of an increase for products from more than one category as subjects progress through the task. Notice that in each case the subject is adopting a simpler, more categorical representation of the products. Therefore, while fatigue should have different directional effects on judgment variance, its effects on the fit of similarity scaling techniques should be similar. Whether brands or categories are involved, similarity scaling should be better able to capture or fit judgments collected later in the task.

The discussion provides a basis for positing

explicit hypotheses regarding the effects of fatigue on similarity judgments. Assuming consumers natural response to fatigue is to rely increasingly on more categorical similarities and differences, we make the following predictions.

H1: The standard deviation in judged similarity should decrease for brand level stimuli and increase for category level stimuli as subjects progress through a similarity rating task.

H2: The fit of a scaling solution should increase as subjects progress through a similarity rating task for both brand and category level stimuli.

We now report on a study that tests these hypotheses.

3. Study design

Pairwise similarity judgments were collected for five separate stimulus sets: soft drinks, candy bars, beverages, snack foods, and lunch products. These stimuli were relevant for the student subjects used to provide data. They also represent two different levels of abstraction or generality: brand level alternatives coming from the same traditional or basic level category (soft drinks and candy bars) and superordinate category alternatives coming from more than one basic level category (beverages, snack foods, lunch products). These stimulus sets are presented in Table 1.

Each stimulus set contained 12 product alternatives requiring subjects to make 66 paired comparison ratings. Each subject rated all 66 product pairs for one of the five stimulus sets. A total of 24, 24, 24, 24, and 27 subjects (total $n = 123$) rated the soft drinks, candy bars, beverages, snack foods, and lunch products respectively. Half of the subjects in each group rated the 66 pairs in one random

Table 1
Stimulus sets

Brand level stimuli	
Soft drinks	Candy bars
Sprite	Three Musketeers
Seven Up	Mars Bar
Diet Sprite	Milky Way
Diet Seven Up	Snickers
Orange Crush	M&M Plain
Diet Orange Crush	M&M Peanut
Coke Classic	Hershey's Plain
New Coke	Hershey's Almond
Pepsi	Nestle's Crunch
Cherry Coke	Reece's Peanut Butter Cups
Diet Coke	Twix Caramel
Diet Pepsi	Kit Kat

Category level stimuli		
Beverages	Snacks	Lunch products
Ice Cream Soda	Popcorn	Carrot
Milk Shake	Nacho Chips	Apple
Chocolate Milk	Crackers	Fruit Juice
Milk	Potato Chips	Yogurt
Fruit Juice	Cheese	Milk
Lemonade	Grapes	Ice Cream
Soft Drink	Apple	Cookie
Diet Soft Drink	Yogurt	Candy Bar
Club Soda	Ice Cream	Soft Drink
Iced Tea	Cookie	Pizza
Bottled Water	Candy Bar	Chicken Sandwich
Iced Coffee	Brownie	Hamburger

order and the other half rated the same 66 pairs in the reverse order. All pairs were rated on an 11-point similarity rating scale ranging from 0 (very dissimilar) to 10 (very similar). Subjects were run through the task in small groups (approximately 20 per group) and were led through the instructions by an experimenter. The instructions included a list of the twelve products the subject would be rating along with a sample similarity judgment scale. Overtly exposing the subjects to the range of products in the task and the similarity scale should help minimize any adaptation to the task. Any major changes in the similarity judgments should, as a result, be due primarily to fatigue.

4. Analysis

The hypotheses were tested by comparing "first half" ratings, based on the subjects' first 33 pairs rated, with "second half" ratings, based on the subjects' second 33 pairs rated. Means and standard deviations were calculated for the first half and second half of each subject's similarity ratings. Each half was then scaled using non-metric multidimensional scaling in both two and three dimensions. Our choice of non-metric MDS was based on several considerations. Non-metric MDS is very commonly used in research applications (Cooper, 1983). Unlike many tree scaling or clustering techniques, non-metric MDS is also applicable to incomplete data matricies (Malhotra, Jain and Pinson, 1988). Finally, non-metric MDS yields solutions that are very similar to those of earlier metric MDS without the solutions being as sensitive to the particular input data (Green, 1975).

Two traditional measures of fit were obtained from both the two- and three-dimensional solutions, monotonic stress and linear R^2 (Kruskal, 1964; Pruzansky, Tversky and Carroll, 1982). Again, if subjects adopt simpler, more categorical representations over time, scaling solutions based on first half data should be more complex and exhibit higher stress than solutions based on second half data. Moreover, judgment variance should decrease for products from the same category and increase for products from different categories.

A series of general linear models were used to test for differences in each of the five dependent variables, the standard deviation of the similarity judgments and the four fit measures. Each model included the two primary independent variables of interest (a two-level variable for first half vs. second half judgments and a two-level brand vs. category level stimulus variable) and their interaction. Hypothesis 1 predicts a significant interaction between these variables on the standard devi-

ation of the judgments. Hypothesis 2 predicts a main effect for first half vs. second half judgments on each of the fit measures. Each analysis model also included random effects (dummy variables) to capture (i) the difference between the two stimulus sets nested under the brand level stimuli (soft drinks and candy bars), (ii) the differences among the three stimulus sets nested under the category level stimuli (beverages, snacks, and lunch products), and (iii) the differences between the two random order conditions nested under each of the five different stimulus sets. [1]

5. Results

We first examine the standard deviation of the judgments presented in Table 2. Overall, judgment variance was significantly greater for the category level stimuli ($F = 25.48$, $p < 0.001$), which is natural given their greater inherent heterogeneity. There was also a main effect for first vs. second half judgments on judgment variance ($F = 6.22$, $p < 0.05$). This main effect is driven, however, by a very significant half by stimulus level interaction ($F = 19.76$, $p < 0.001$). As predicted by Hypothesis 1, judgment variance decreased as the subjects progressed through the brand level stimuli and increased as they progressed through the category level stimuli. This result supports the notion that subjects adopt simpler representations and rely more exclusively on categorical similarities and differences as they progress through a similarity judgment task. The observed interaction is inconsistent

---

[1] To provide a finer grade analysis, the judgments were also broken into thirds and examined. That is, we contrasted the first 22 pairs rated, the second 22 pairs rated, and the third and last 22 pairs rated. The results of the three-way data analyses were very consistent with the two-way (first half vs. second half) analyses. Moreover, the fit of the two-way analytical models dominated the fit of the three-way models. We shall concentrate, therefore, on the two-way, first half vs. second half judgment data.

Table 2

Differences between first and second half paired comparison similarities

Brand level stimuli					
Stimuli	Standard deviation	2-D stress	3-D stress	2-D R^2	3-D R^2
Soft drinks:					
First half	2.840	0.100	0.027	0.465	0.714
Second half	2.094	0.043	0.006	0.637	0.816
Candy bars:					
First half	2.135	0.112	0.043	0.452	0.639
Second half	1.892	0.038	0.005	0.659	0.814
Category level stimuli					
Stimuli	Standard deviation	2-D stress	3-D stress	2-D R^2	3-D R^2
Beverages:					
First half	2.380	0.109	0.035	0.573	0.758
Second half	2.616	0.057	0.010	0.680	0.864
Snacks:					
First half	2.664	0.091	0.036	0.582	0.733
Second half	2.688	0.063	0.010	0.655	0.859
Lunch products:					
First half	2.531	0.098	0.036	0.533	0.676
Second half	2.686	0.049	0.010	0.605	0.797

with the notion that subjects simply adopt a particular response level or became careless as they proceed. [2]

The analyses of the fit measures provide additional support for the notion that representations simplify through the course of a similarity judgment task. The stress (badness-of-fit) and R^2 (goodness-of-fit) measures for both the two- and three-dimensional MDS solutions are presented in Table 2. There were very significant decreases in stress for both the two-dimensional solutions ($F = 138.776$,

$p < 0.0001$) and three-dimensional solutions ($F = 150.036$, $p < 0.0001$) from the first to the second half judgments. The average two-dimensional stress was 0.102 and 0.050 respectively for the first and second half data. The corresponding averages for three-dimensional stress were 0.035 and 0.008, respectively. There was no difference in stress for brands vs. categories in either case.

There was, however, a significant half by level interaction effect on two-dimensional stress ($F = 6.016$, $p < 0.05$). The decrease in stress from first to second half judgments was slightly greater for the brands than for the categories. This draws into question the interpretability of the main effect difference between the first and second half similarity judgments. As a check, first vs. second half judgments were examined within each of the five stimulus sets. Consistent with the main effect predicted under Hypothesis 2, each of the five stimulus sets exhibited a significant ($p < 0.05$) decrease in stress from first to second half judgments.

The linear fit results parallel those for stress. As predicted, there was a significant increase in R^2 for both the two-dimensional ($F = 74.310$, $p < 0.0001$) and three-dimensional ($F = 107.793$, $p < 0.0001$) solutions. The average R^2's were 0.521 and 0.646 respectively for the first and second half two-dimensional solutions, and 0.704 and 0.829 respectively for the first and second half three-dimensional solutions. There was again a significant half by stimulus level interaction effect on R^2 in the case of two-dimensional solutions ($F = 11.139$, $p < 0.001$). The increase in linear fit from the first to the second half judgments was greater for the brand than the category level stimuli. However, separate analyses again revealed significant ($p < 0.05$) increases in two-dimensional R^2 within each of the five stimulus sets.

Our analysis thus supports the predicted general increase in fit from first to second half judgments. At the same time, the ob-

[2] Though no predictions were made, the possible effects of stimulus level and first vs. second half judgments on mean similarity were also examined. There were no significant main effects on mean similarity for either variable. There was a marginally significant half by level interaction ($F = 2.9$, $p < 0.10$). The brand stimuli became slightly less similar and the category stimuli became slightly more similar over time. However, there were no differences in average first half vs. second half judgments within any of the five individual stimulus sets.

served interactions involving the first vs. second half judgments and the brand vs. category level of the stimuli on the two-dimensional fit measures were unexpected. In hindsight, these results are likely due to the salient across category differences that pervade the category level stimuli. These categorical differences should, for the most part, drive both the first half and second half judgments of the across category stimuli. In contrast, the predicted shift in focus from within category differences to common category membership might explain the more pronounced effects on fit for the brands.

A difference between the stress and R^2 results was also observed. Unlike the stress measures, the R^2's for the category level stimuli were overall greater than those for the brand level stimuli for both the two-dimensional ($F = 9.785$, $p < 0.005$) and three-dimensional ($F = 8.494$, $p < 0.005$) solutions. The existence of clear common and distinctive features for brand level stimuli (Johnson and Fornell, 1987) may explain this result. These features resulted in a greater clustering of stimuli at the brand level. This clustering resulted in more pronounced non-linear relationships between the original and fitted distances for the brands.

The results found can be shown to affect the interpretation and content of scaling solutions. To illustrate, first vs. second half judgments for the soft drink stimuli were pooled across subjects and scaled. The two-dimensional MDS solutions for the first vs. second half data are shown in the Fig. 1 (Kruskal's stress equalled 0.144 and 0.100, respectively). Notice that certain differences among the brands affected the first half judgments more than the second half judgments. In particular, there is a clearly interpretable diet vs. non-diet dimension in the first half data which does not materialize for the second half data. The brands are clustered primarily on flavor in the solution based on second half judgments. The two solutions suggest different competi-
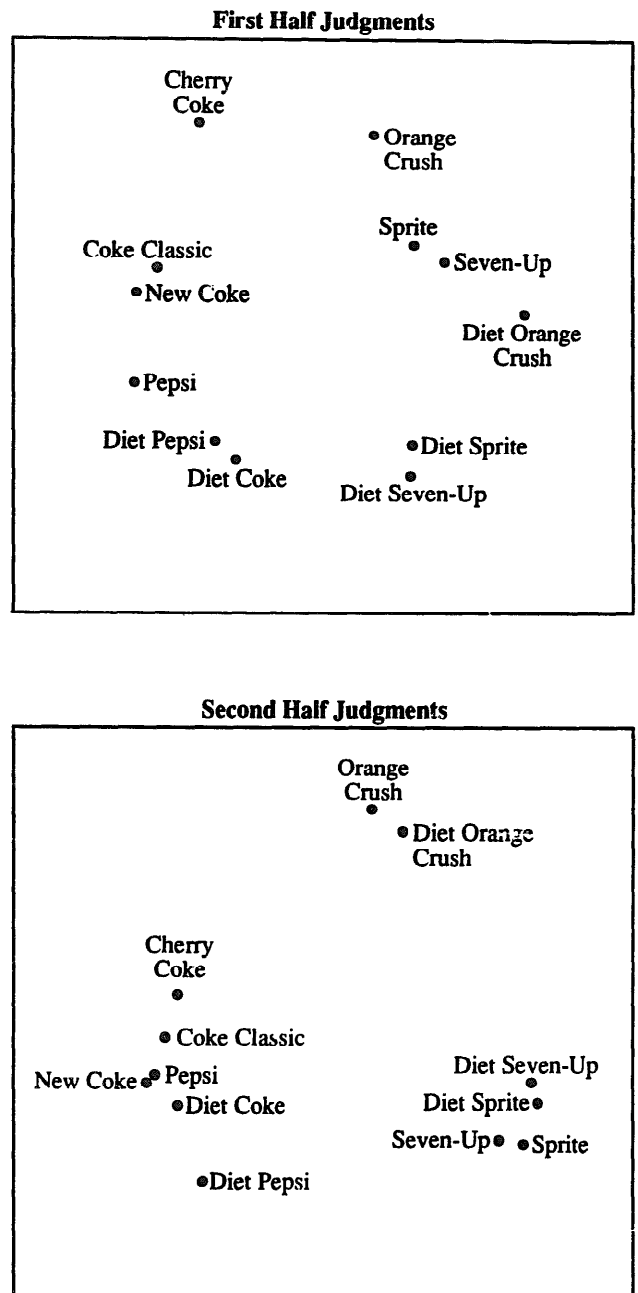

Fig. 1. Two-dimensional MDS solutions for first vs. second half soft drink judgments.

tive situations and underscore the practical importance of the results.

6. Conclusions

Overall, the results support Hypotheses 1 and 2 and demonstrate the systematic effects

of fatigue on ratings of product similarity. Judgment variance increased for the category level stimuli while it decreased for the brand level stimuli through the course of a similarity judgment task. The fit of MDS solutions was also greater for judgments collected later in the task than for judgments collected early. These results support a general reduction in the complexity of respondents' product representations and judgments over time.

The results are very consistent with a categorization-based explanation of the effects of fatigue on similarity judgments. At least for the stimuli and rating task in the present experiment, it appears that consumers relied increasingly on similarities and differences in product category membership as a means of adapting to fatigue. Naturally, the results may depend on the particular stimuli used in the study (non-durable food products). Any general conclusions regarding the effects of fatigue on perceptions and judgments awaits further research involving qualitatively different stimuli and rating tasks.

The results do, however, suggest important implications for marketing research. At least for the task studied here in which subjects provided 66 paired product comparison ratings, it appears that judgments simplify rather than degenerate over time. Our subjects appeared to rely increasingly on categorical product similarities and differences. They did not simply adopt a restricted response set or become careless. As shown in Fig. 1, these fatigue effects may have a significant impact on the interpretation of scaling output. In light of the results, researchers might explore newer probabilistic MDS models that allow for inherent variance in the input data (MacKay, 1989).

The study also suggests a number of important directions for future research. One is to explore fatigue effects as a function of the number of judgments involved in a similarity rating task. Another is to explore potential differences in fatigue effects across individu-

als. Finally, the apparent importance of product categorization in consumer perceptions underscores the need for research on alternative data collection procedures. In particular, relatively little attention has been paid to the use of sorting or grouping tasks in marketing applications involving larger stimulus sets (see Rao and Katz (1971) for a notable exception). Product sorting tasks are very consistent with a categorization perspective and would appear to deserve considerable attention.

References

Arabie, P., J.D. Carroll, W.S. DeSarbo and J. Wind, 1981. Overlapping clustering: A new method for product positioning. Journal of Marketing Research 18, 310–317.

Cooper, L.G., 1983. A review of multidimensional scaling in marketing research. Applied Psychological Measurement 7, 427–450.

David, H.A., 1988. The method of paired comparisons (2nd ed.). London: Charles Griffen.

Day, G.S., T. Deutscher and A.B. Ryans, 1976. Data quality, level of aggregation and non-metric multidimensional scaling solution(s). Journal of Marketing Research 13, 92–97.

Dong, H., 1983. Method of complete triads: An investigation of unreliability in multidimensional perception of nations. Multivariate Behavioral Research 18, 85–96.

Green, P.E., 1975. Marketing applications of MDS: Assessment and outlook. Journal of Marketing 39, 24–31.

Green, P.E. and F.J. Carmone, 1970. Multidimensional scaling and related techniques in marketing analysis. Boston, MA: Allyn & Bacon.

Green, P.E. and V.R. Rao, 1972. Applied multidimensional scaling: A comparison of approaches and algorithms. New York: Holt, Rinehart & Winston.

Hauser, J.R. and F.S. Koppelman, 1979. Alternative perceptual mapping techniques: Relative accuracy and usefulness. Journal of Marketing Research 16, 495–506.

Jain, A.K. and C. Pinson, 1976. The effects of order of presentation of similarity judgments on multidimensional scaling results: An empirical examination. Journal of Marketing Research 13, 435–439.

Johnson, M.D. and C. Fornell, 1987. The nature and methodological implications of the cognitive representation of products. Journal of Consumer Research 14, 214–228.

Johnson, S.C., 1967. Hierarchical clustering schemes. Psychometrika 32, 241–254.

Kendall, M.G., 1955. Further contributions to the theory of paired comparisons. Biometrics 11, 43–62.

Klahr, D., 1969. A Monte Carlo investigation of the statistical significance of Kruskal's non-metric scaling procedure. Psychometrika 34, 319–330.

Kruskal, J.B., 1964. Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. Psychometrika 29, 1–27.

MacKay, D.B., 1989. Probabilistic multidimensional scaling: An 'nisotropic model for distance judgments. Journal of Mathematical Psychology 33, 187–205.

Malhotra, N.K., 1987. Validity and structural reliability of multidimensional scaling results. Journal of Marketing Research 24, 164–173.

Malhotra, N.K., A.K. Jain and C. Pinson, 1988. The robustness of MDS configurations in the case of incomplete data. Journal of Marketing Research 25, 95–102.

Medin, D.L. and E.E. Smith, 1984. Concepts and concept formation. Annual Review of Psychology 35, 113–138.

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

Moore, W.L. and D.R. Lehmann, 1982. Effects of usage and name on perceptions of new products. Marketing Science 1, 351–370.

Murphy, G.L., 1982. Cue validity and levels of categorization. Psychological Bulletin 91, 174–177.

Murphy, G.L. and E.E. Smith, 1982. Basic-level superiority in picture categorization. Journal of Verbal Learning and Verbal Behavior 21, 1–20.

Pruzansky, S., A. Tversky and J.D. Carroll, 1982. Spatial vs. tree representations of proximity data. Psychometrika 47, 3–24.

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

Rosch, E., 1975a. Cognitive reference points. Cognitive Psychology 7, 532–547.

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

Rosch, E., C.B. Mervis, W.D. Gray, D.M. Johnson and P. Boyes-Braem, 1976. Basic objects in natural categories. Cognitive Psychology 8, 382–439.

Sattath, S. and A. Tversky, 1977. Additive similarity trees. Psychometrika 42, 319–345.

Shepard, R.N., 1962. The analysis of proximities: Multidimensional scaling with an unknown distance function. Parts I and II. Psychometrika 27, 125–140 and 219–246.

Shepard, R.N. and P. Arabie, 1979. Additive clustering: Representation of similarities as combinations of discrete overlapping properties. Psychological Review 86, 87–123.

Srivastava, R.K., R.P. Leone and A.D. Shocker, 1981. Market structure analysis: Hierarchical clustering of products based on substitution-in-use. Journal of Marketing 45, 38–48.

Sudman, S. and N. Bradburn, 1982. Asking questions. San Francisco: Jossey–Bass.

Summers, J.O. and D.B. MacKay, 1976. On the validity and reliability of direct similarity judgments. Journal of Marketing Research 13, 289–295.

Tversky, A., 1977. Features of similarity. Psychological Review 84, 327–352.

Weksel, W. and E.E. Ware, 1967. The reliability and consistency of complex personality judgments. Multivariate Behavioral Research 2, 537–541.