





UNIVERSITY OF  
ILLINOIS LIBRARY  
AT URBANA-CHAMPAIGN  
BOOKSTACKS

**CENTRAL CIRCULATION BOOKSTACKS**

The person charging this material is responsible for its renewal or its return to the library from which it was borrowed on or before the **Latest Date** stamped below. **You may be charged a minimum fee of \$75.00 for each lost book.**

**Theft, mutilation, and underlining of books are reasons for disciplinary action and may result in dismissal from the University.**

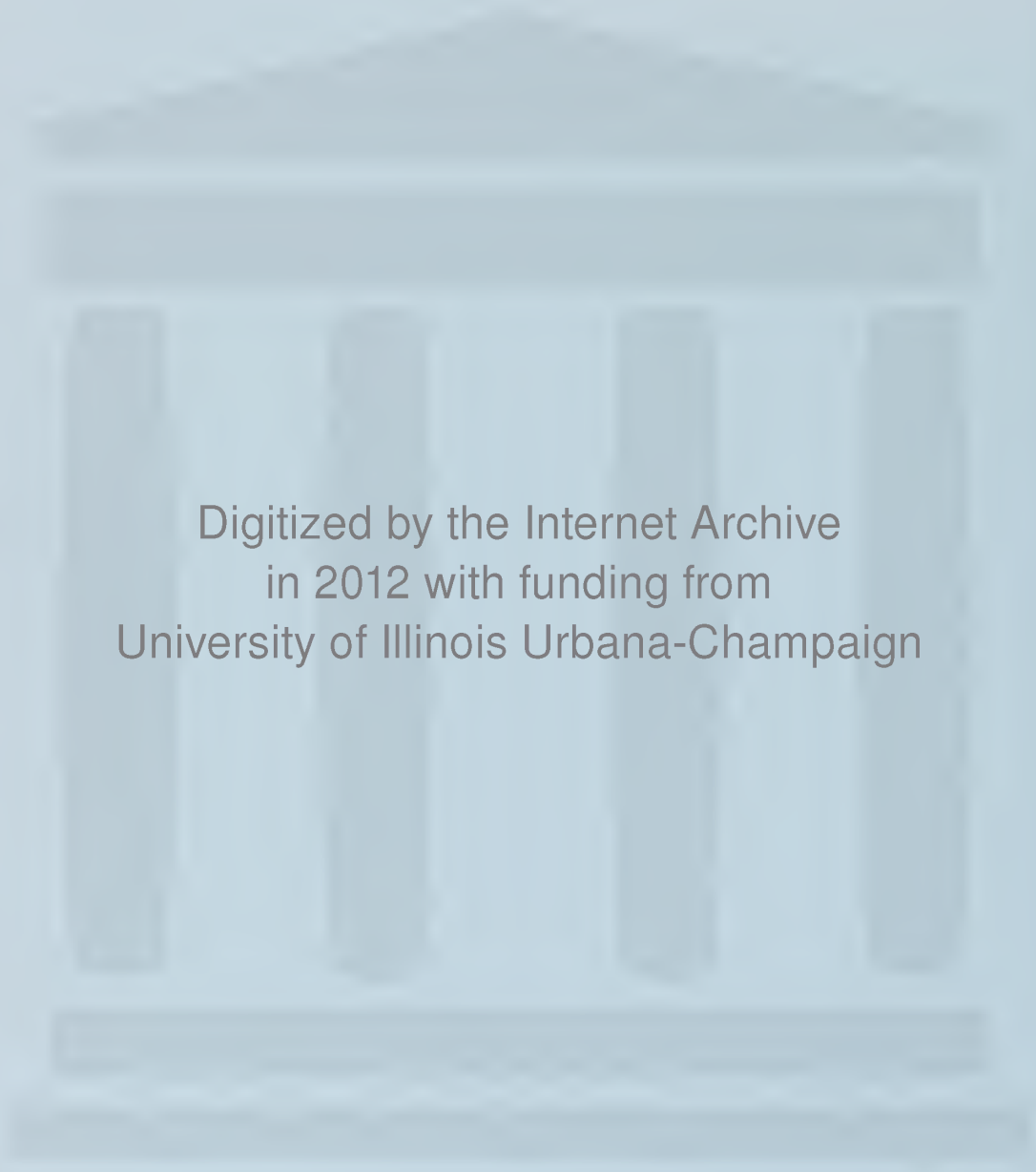
**TO RENEW CALL TELEPHONE CENTER, 333-8400**

**UNIVERSITY OF ILLINOIS LIBRARY AT URBANA-CHAMPAIGN**

AUG 01 1997

When renewing by phone, write new due date below previous due date.

L162



Digitized by the Internet Archive  
in 2012 with funding from  
University of Illinois Urbana-Champaign

<http://www.archive.org/details/responsecategori92108view>

330  
B385  
1992:108 COPY 2

STX

# Response Categories as Fuzzy Sets: A Fuzzy Set Theoretic Perspective on Issues in Scale Development

The Library of the

APR 1 1992

University of Illinois  
at Urbana-Champaign

**Madhubalan Viswanathan**  
*Department of Business Administration*  
*University of Illinois*

**Shantanu Dutta**  
*Department of Marketing*  
*University of Chicago*

**Mark Bergen**  
*Department of Marketing*  
*University of Chicago*

**Terry Childers**  
*Department of Marketing*  
*Carlson School of Management*



# BEBR

FACULTY WORKING PAPER NO. 92-0108

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

February 1992

Response Categories as Fuzzy Sets: A Fuzzy Set Theoretic Perspective on Issues in Scale  
Development

Madhubalan Viswanathan\*

Mark Bergen

Shantanu Dutta

and

Terry Childers

\*Madhubalan Viswanathan is Assistant Professor in the Department of Business Administration at the University of Illinois, Champaign, IL 61820. Mark Bergen and Shantanu Dutta are Assistant Professors in the Department of Marketing, School of Business, University of Chicago, Chicago, IL. Terry Childers is Associate Professor in the Department of Marketing, Carlson School of Management, University of Minnesota, Minneapolis, MN 55455.





## Abstract

This paper suggests the application of the concept of fuzzy sets to issues relating to scale development. Specifically, response categories of a scale are conceptualized as fuzzy sets (i.e., sets whose members can have varying degrees of membership in it rather than either belong or not belong to it) and the issue of the optimal number of response categories to use in a scale is examined while some other issues in scale development are also discussed. Rather than allow the choice of a single response category as a response as is the case with traditional scales, a new type of scale is proposed which allows for the choice of one or more response categories with the assignment of membership values to each response category. Norms are suggested for the use of this scale during the development phase of traditional scales. A study is reported where responses to stimulus-centered, response-centered, and behavioral frequency items were collected using this new scale while manipulating the number of response categories in the scale across groups of subjects. The results are interpreted in terms of recommendations for the choice of the optimal number of response categories. Other possible applications of this conceptualization are also discussed.



This paper suggests the application of concepts from fuzzy set theory to issues relating to scale development. Specifically, response categories of a scale are conceptualized as fuzzy sets and the issue of the optimal number of response categories to use in a scale is examined. A fuzzy set, as distinct from a crisp set, is one whose members can have varying degrees of membership in it rather than either belong or not belong to it (Zadeh, 1976). Therefore, a fuzzy set allows for different degrees of membership in it. The notion of degrees of membership as suggested in the context of fuzzy sets has been used to understand the gradedness of membership of instances in natural categories (cf. McCloskey and Glucksberg 1978). It is suggested here that scale responses could belong to more than one response category with different degrees of membership. In contrast to traditional scales which require the choice of a single response category as a response, a new scale is proposed which allows for responses that indicate degrees of membership in one or more response categories. A study using this scale is reported where data was collected for stimulus-centered, response-centered, and behavioral frequency items, using different number of response categories across groups of subjects. Implications of the new scale in providing diagnostic information during scale development as well as other possible applications of the proposed conceptualization are discussed.

The rest of the paper is organized as follows. The notion of a fuzzy set is described briefly. A discussion wherein response categories are conceptualized as fuzzy sets is presented with a view to bringing out possible applications of the notion of fuzzy sets to scale development issues. Specifically, research on the optimal number of response categories is reviewed and the conceptualization of response categories as fuzzy sets is applied to this problem. A study is reported which used a new type of scale to capture the notion that response categories can be viewed as fuzzy sets. Finally, several applications of the proposed conceptualization are discussed.

### RESPONSE CATEGORIES OF A SCALE AS FUZZY SETS

This section suggests the conceptualization of response categories of a scale as fuzzy sets as a means to addressing issues such as the optimal number of response categories to use in a scale. The notion of a fuzzy set is described and it is suggested that response categories are similar to fuzzy sets. Insights drawn from this conceptualization for the issue of the number of response categories

to use in a scale are discussed.

### Introduction to Fuzzy Sets

Zadeh (1976) suggested the notion of a fuzzy set as distinct from a crisp set. The notion of a fuzzy set has been used to explain several phenomena such as membership of instances in natural categories. Zadeh's explanation of the nature of fuzzy sets can be understood using an example involving scale response. Consider a scale to measure ratings of gas mileage of automobiles using three response categories, 'high', 'medium', and 'low'. Say, respondents are aware of gas mileage of automobiles to the nearest mpg and are rating automobiles with gas mileage ranging from 20 to 40 mpg. Considering their responses with respect to the response category 'high', most respondents may consider 20 mpg as definitely not being 'high' mileage but definitely being 'low' mileage and 40 mpg as definitely being 'high' mileage. Similarly, many consumers may consider 25 mpg as definitely not being 'high' mileage. However, a certain number of mpgs. above 25 mpg could be considered as 'high' mileage. This raises the question as to when the transition from 'not high' (i.e., 'low' or 'medium') to 'high' occurs. If an arbitrary criterion is set such that any mileage which is one mpg greater than 30 is considered 'high' then the distinction between 'high' and 'not high' (i.e., 'medium' or 'low') reduces to being equivalent to one mpg. This raises the issue as to where a magnitude such as 30.5 mpg would belong. Criteria could be set to suggest even smaller values of mileage as distinguishing 'high' and 'not high'. However, the use of an arbitrary criterion to define an inherently imprecise category leads to minute distinctions between 'high' and 'not high'. If large intervals such five mpg are used to set a criteria, then the intermediate range of magnitudes (from 30 to 35) is undefined. A definition of 'high' as being 1 mpg higher than any other mileage that is considered 'high' would result in all mpg being considered as 'high' mileage.

Zadeh (1976) attempts to resolve this paradox using the notion of fuzzy sets. Applying Zadeh's explanation to the present example, terms such as 'high' are vague or imprecise and there is a gradual transition from mpgs. that are 'not high' to mpgs. that are 'high' mileage. A category such as 'high' is called a fuzzy set (as opposed to a crisp set) since it eliminates the sharp distinction between members and nonmembers and allows for grades of membership. A fuzzy set is defined in mathematical terms by assigning a degree of membership to each instance or member to indicate its degree of membership in the set. In the present example, each mileage could be given a value



representing its degree of membership in the category 'high', with higher membership values representing greater degrees of membership. Similarly, each mileage could be given membership values representing degrees of membership in the categories 'medium', and 'low'.

### Response Categories as Fuzzy Sets

While the example above relates to the single category, 'high', a similar line of reasoning can be extended to a set of response categories in a scale. This is the case of a categorical scale that is typically used in measurement. A group of response categories or fuzzy sets are used to capture responses along some continuum. Therefore, responses may be analyzed in terms of membership (i.e., non-zero degrees of membership) in one or more of these response categories rather than perfect membership in a single category. Traditional scales, by requiring the choice of a single response category, implicitly assume that responses have perfect membership in a single response category. The use of categorical scale anchors in combination with the requirement for the choice of a single category as a response potentially leads to loss of information about degrees of membership in more than one response category.

Considering the mileage example, a range of mileages could belong to the category 'high' with different degrees of membership. For example, 32 mpg may be considered as belonging to the category 'high' with a membership of 0.5 while 30 mpg may have a membership of 0.4. Further, 32 mpg may also belong to the category 'medium' with a membership of 0.2. The key point to note is that response categories are inherently fuzzy or imprecise in nature and that several responses may be partial or complete members in one or more categories. Therefore, the argument advanced here is that response categories are similar to natural categories in terms of allowing graded membership in them (cf. Rosch 1973). Gradedness in natural categories has been argued to occur due to various combinations of featural and dimensional values leading to a continuum of membership in a category. Graded membership of responses in response categories is argued to occur due to the use of imprecise response categories to represent a continuum.

### Applications to the Issue of the Number of Response Categories in a Scale

Viewing response categories as fuzzy sets, insights can be gained about the optimal number of response categories to utilize in a scale. Several researchers in the past have addressed the problem of assessing the optimal number of response categories to employ in a scale. Cox (1980), in

reviewing the literature in this area of research, points out that suggestions made by researchers range from the use of two to 25 alternatives. Approaches in the past include assessment of psychometric properties of scales with different number of response categories, the use of approximately seven response categories based on research on absolute judgments, and the information theoretic approach to determine information transmitted by a scale (Cox 1980). While seven levels of magnitude are often cited as being ideal for measurement scales since human ability is assumed to lie in the vicinity of this number, Cox (1980) points out that this rule was derived from findings in the theoretical context of absolute judgments on perceptual stimuli (Miller, 1954) and may not be generalizable to issues concerning long term memory. It is possible that human ability to discriminate and provide responses may vary widely as a function of factors such as individual expertise in a domain and the nature of dimensions being measured, thereby necessitating the tailoring of scales to various situations. Cox (1980) suggests that there is an immediate need to develop methods at the pretesting stage to evaluate the nature of information being collected using different number of response categories. This is argued to be the case, particularly for stimulus-centered scales, since response centered scales involve use of multiple items which increases the effective redundancy of information and the effective variance of the scale (Cox 1980).

The nature of trade-offs involved in increasing the number of scale points in a measurement scale have been discussed in the past (Cox, 1980). It has been suggested that as the number of scale points are increased, there is an increase in information that is transmitted along with a possible increase in response error. This error occurs due to consumers' cognitive limitations for using a large number of scale points. The use of categorical labels (such as 'high' instead of say, 32 mpg) to capture responses involves the reduction in resolution which is compatible with human abilities and reduces this type of response error.

It will be argued that non-zero membership of responses in multiple response categories may arise when there is a mismatch between responses and response categories in terms of their precision or fine-grainedness.<sup>1</sup> Two possible scenarios will be considered wherein responses are more precise and less precise than response categories. Consider a scenario where scale responses are more discriminating than response scales used to measure them as was the case with the automobile example discussed above. Therefore, relatively precise or fine-grained responses have to be reduced

to fit a set of relatively imprecise response categories. As a result, no single response category may completely capture a response. Rather, the response may have varying degrees of membership in more than one response category. In Figure 1, the response of 27 mpg does not fit completely into any response category but overlaps with two categories to different degrees. Note that whenever responses are more precise or fine-grained than response categories (i.e., involve the use of more categories to describe a continuum than the response scale), the possibility that a single response category does not completely capture a response arises. Such responses are due to the use of categorical or imprecise labels to represent a continuum, thereby leading to the possibility of graded membership of responses in one or more of these categories.

-----  
Insert Figure 1 about here  
-----

It may seem that the mismatch stated above could be resolved if response scales are then designed to be at least as precise as responses. However, a similar problem exists if response categories are more precise than responses (see Figure 1). A relatively imprecise response such as 'above average' mileage overlaps with two categories on the response scale (i.e., high and very high) therefore leading to the possibility of membership in each of these two categories. The problem here is the reverse, to match relatively imprecise categories to a more fine-grained scale. Therefore, more than one response category may be chosen for any particular response. However, the traditional requirement of the choice of a single response category restricts the responses to a single category. Hence, as long as there is a mismatch in terms of the number of scale points in the scale versus memory, there is loss of information due to the requirement for a single category response.<sup>2</sup>

Given the nature of responses that may arise as a function of the number of response categories issue, responses collected on a scale that allows multiple responses and varying degrees of membership could provide important diagnostic information on the number of response categories issue. By varying the number of categories on such a scale and studying the extent to which more than one response category is utilized by respondents for a set of items, valuable information may be collected about the optimal number of response categories to use in a particular situation. Ideally, to



the extent that respondents tend to use a single response category with perfect membership to characterize their response, use of an appropriate number of response categories is suggested. Extending this argument, scales with different number of response categories can be compared to assess the extent to which single response categories with high levels of membership are used. As responses approach the “ideal” described above, the number of response categories used could be argued to be more and more appropriate. Therefore, responses to such scales provide a basis to choose between scales with different number of response categories.

#### Other Factors in Scale Development

The discussion to this point has focused on the issue of the number of response categories. However, several other factors may also lead to the need for a scale that allows for the type of responses described above and two such factors are discussed briefly. A subtle type of error occurs when there is a mismatch in terms of descriptors used to label response categories. Consider a scenario where a behavioral frequency item has a scale whose response categories are completely described (e.g., for an item on frequency of visit to malls, a set of labels such as ‘once a year’, ‘once a month’, and ‘once a week’). To the extent that the set of labels do not match the responses provided by respondents, membership of responses in more than one category may occur. A respondent who visits the mall once in two weeks may have to choose both ‘once a month’ and ‘once a week’ with some level of membership in each. Such responses with membership in multiple response categories arise due to a mismatch between the set of descriptors used in a scale and the responses provided. Again, by varying the descriptors on such a scale and studying the extent to which more than one response category is utilized by respondents for an item (or a set of items) valuable information may be collected about the descriptors to use to label response categories in a particular situation.

The scenarios described above relate to mismatches between responses and response categories in terms of the number of response categories or between responses and descriptors of specific response categories. It is also possible that some responses *inherently* involve more than one response category due to some sort of aggregation across situations or time that is required to provide a response. Consider a behavioral frequency item as described above that requires an estimate of the frequency of visit to a mall. If a respondent usually visits a mall once a month but



sometimes visits it once in two weeks, the response would have some degree of membership in both these categories. This represents a scenario where the response inherently involves multiple categories, *irrespective of how precise the categories are or how they are labeled*. Therefore, such responses cannot be captured by the appropriate number of response categories and/or category descriptors. Similarly, consider a response to a response-centered item such as “I am an intellectual” with response categories from Strongly disagree to Strongly agree. Again, to the extent that some form of aggregation across, perhaps, the various roles played by the individual which relate to this item is required, a response may be a member of more than one response category. (Such aggregation may be more likely to occur to the extent that an item is general and not specific, since general items may require aggregation across specific situations). Such information cannot be collected completely using traditional scales but may be critical for input to further analyses. It represents the spread or range of an individual’s response to an item. The incorporation of such information may explain a portion of the unexplained error in predictive models as well as the study of relationships using other statistical analyses.

## METHOD

This section suggests the use of a scale that assesses the fuzziness of response categories by allowing responses that can belong to multiple response categories with different degrees of membership. This scale is derived from past research (Smithson 1982) which used a fuzzy theoretic framework to develop techniques for coding qualitative data. In coding tasks involving the classification of observation into sets predetermined by categories, observations may not precisely fit a simple category. Researchers have suggested the use of certain phrases to indicate degrees of membership of items in categories (Lakoff 1973; Kempton 1978). Using a range of phrases suggested by Kempton (1978), Smithson suggests the assignment of membership values to items to indicate their memberships to various categories. The suggested phrases and membership values are as follows: "completely described by the coding scheme," "mostly described by the coding scheme," "sort of described by the coding scheme," "not too well described the coding scheme," "not really described by the coding scheme," and "not at all described by the coding scheme," with suggested membership values of 1.0, 0.8, 0.6, 0.4, 0.2, and 0.0, respectively (Smithson 1982).

As Smithson (1982) points out, an item could have different degrees of membership in more than one set. Conceptualizing response categories as fuzzy sets, a response could have differing degrees of membership in response categories. A new type of scale was used here that was derived from suggestions in past research (Smithson 1982; Kempton 1978). This scale allowed respondents to assign degrees of membership to each response category to indicate the extent to which a response was captured by that category (see Appendix for an example of the scale with instructions). The levels of membership and the phrases suggested by Smithson (1982) were used with the replacement of the phrase “coding scheme” with the word “alternatives”. A variation of the scale which required respondents to write in membership values was pilot tested and the scale was modified such that respondents could perform the easier task of circling membership values. Detailed instructions on the use of the scale and several sample trials were provided to ensure appropriate utilization of this new scale. The description of each membership level was repeated at the top of each page of the questionnaire administered to collect data. Several self-report measures relating to respondents reactions to the use of this scale were collected during the pilot test and the study.

-----  
 Insert Figure 2 about here  
 -----

## Overview and Procedure

The approach taken here was to collect responses for a range of different items across groups that were assigned to conditions with different number of response categories. Three groups of 30 subjects at a midwestern university were assigned to conditions where 3, 5, and 7 response categories, respectively, were used for scales. Hence, the number of response categories used in scales were manipulated between groups of subjects using three levels. Data was collected on three types of items, stimulus centered, response-centered, and behavioral frequency items using a questionnaire. Responses to stimulus centered items involved rating how much respondents liked a set of twelve soft drinks on scales that were end-anchored Very Bad-Very Good. Responses to response-centered items involved the use of an 16 item version of the Need for Cognition scale (Perri and Wolfgang 1988) using scales that were end-anchored Strongly Disagree-Strongly Agree. Behavioral frequency items involved responses to two items on hours of daily television viewing and frequency of visits to the movie theater. These scales were completely described with a range of

response categories.

Subjects were provided with detailed instructions to complete scales and completed several sample trials. The instructions followed Smithson (1982) in describing the use of various response categories. Further, the membership values and their description were presented on the top of every page of the questionnaire. Responses required subjects to circle a set of values for each response category to indicate membership of the response in that response category. Non-response to a response category indicated a membership value of 0.0. After filling out these scales, subjects filled out scales pertaining to their reactions to the use of the new scale.

### Data Analysis and Results

Several scales were used to assess subjects' reactions to using the new scale. Mean ratings across all 90 subjects for these items appeared to be satisfactory and are as follows; motivation to complete scales (10 point scale anchored Not at all motivated -Very motivated; 6.33/10), knowledge level to complete scales (10 point scale anchored Very low-Very high; 7.47/10), familiarity with completing scales (10 point scale anchored Very low-Very high; 5.76/10), adherence to instructions (10 point scale anchored To a large extent-Not at all; 4.43/10), confidence in responses provided (10 point scale anchored Very low-Very high; 6.99/10), satisfaction with accuracy of responses (10 point scale anchored Very dissatisfied-Very satisfied; 6.88/10), certainty in responses (10 point scale anchored Not at all certain-Very certain; 6.78/10), sureness in responses (10 point scale anchored Not at all sure-Very sure; 6.86/10), and ease of filling scales (10 point scale anchored Very difficult-Very easy; 6.32/10). These results suggest that the new scale was completed with moderately high levels of motivation, adherence to instructions, and knowledge. Further, moderately high ratings of confidence, certainty, and perceived accuracy, in the responses provided were also found.

Using the norm discussed earlier that responses belonging to a single category with a membership of 1.0 represented an ideal scenario since traditional scales allowed only such a response, several indicators of distance from this "ideal" were computed for scale response data. Therefore, these indicators were measures of the extent to which a single category captured the response for a scale completely. One indicator was the maximum membership value that was assigned to any of the response categories of a scale. This was on the basis that a high membership value for a response category on a scale indicated less overlap between response categories. Ideally,



a membership value of 1.0 suggests that a response is completely described by the response category. Therefore, higher maximum values are indicative of more appropriate number of response categories since they suggest that a scale is closer to the ideal of a membership value of 1.0 for a particular response category. Relatively low maximum membership values values are indicative of lower membership in a particular response category. Therefore, it is suggested that no single response category completely captures a response.

Another indicator, referred to as a fuzzy index, is the difference between the maximum value assigned to a particular response category and the values assigned to the other response categories. If one response category completely captures a response (i.e.,  $m = 1$ ), then this difference would be 1. If several response categories are required to capture a response, this the fuzzy index may be close to 0 or even have negative value. Correlations between the maximum value and the fuzzy index for each set of items for each group of subjects were found to be positive and significant.<sup>3</sup>

#### Results for Stimulus-centered Items

The mean maximum values (MAX), and fuzzy values (FUZZY) were computed for each item for each condition with respect to the number of response categories in a scale. Further, the mean of these indicators across the set of twelve items are also presented. These results are presented in Table 1. As evident from the overall mean and the means for several items, the values of MAX (0.67, 0.72, and 0.76, respectively for 3, 5, and 7 categories) and FUZZY (0.60, 0.62, and 0.67, respectively for 3, 5, and 7 categories) increase with an increase in the number of response categories. Comparisons of MAX values across groups suggested that the 5 category group was marginally higher than the 3 category group ( $t(56) = 1.32$ ;  $p < .10$ ), the 7 category group was directionally higher than the 5 category group ( $t(57) = 1.18$ ;  $p > .10$ ), and 7 category group was significantly higher than the 3 category group ( $t(57) = 2.35$ ;  $p < .05$ ). No significant differences were obtained for comparisons of FUZZY values across groups. It appears based on these results that 7 response categories may be the most appropriate among the three options considered. Speculating on the pattern of results, it is possible that scale with more than 7 categories may perform better than any of these three options. This is argued to be the case since it is possible that responses to items (i.e., degree of liking which is an overall global judgment) may be more discriminating or fine-grained than any of the three options considered.



-----  
 Insert Table 1 about here  
 -----

### Results for Response-centered Items

The results for response-centered items are presented in Table 2. The values of MAX (0.65, 0.73, and 0.82, respectively for 3, 5, and 7 categories) and FUZZY (0.57, 0.62, and 0.69, respectively for 3, 5, and 7 categories) suggest an increase with increase in the number of response categories. Comparisons of MAX values across groups suggested that the 5 category group was significantly higher than the 3 category group ( $t(56) = 2.60; p < .01$ ), the 7 category group was significantly higher than the 5 category group ( $t(57) = 2.77; p < .01$ ), and 7 category group was significantly higher than the 3 category group ( $t(57) = 5.48; p < .01$ ). Comparisons of FUZZY values across groups suggested that the 5 category group was marginally, significantly higher than the 3 category group ( $t(56) = 1.54; p < .10$ ), the 7 category group was directionally higher than the 5 category group ( $t(57) = 1.24; p > .10$ ), and 7 category group was significantly higher than the 3 category group ( $t(57) = 2.50; p < .01$ ). These results suggest that the 7 category scale is the most appropriate of the three options considered. A pattern is observed wherein the indicators provide better values with an increase in the number of response categories. Therefore, it is possible that a scale with more than 7 response categories may be more appropriate than a 7 category scale. Again, this is argued to be the case since it is possible that responses to items (i.e., degree of liking which is an overall global judgment) may be more discriminating or fine-grained than any of the three options considered. On the other hand, if responses involve some degree of spread due to the notion of aggregation discussed earlier, then an improvement may not be observed with an increase in the number of response categories.

-----  
 Insert Table 2 about here  
 -----

### Results for Behavioral Frequency Items

The results for these items are presented in Table 3. These results should be interpreted in light of both the number of response categories used and the specific frequency labels used for each response category since these scales were completely described. For the item on hours of television

viewing, the values of MAX (0.70, 0.82, and 0.79, respectively for 3, 5, and 7 categories) and FUZZY (0.60, 0.69, and 0.71, respectively for 3, 5, and 7 categories) suggest that both the 5 and 7 category scale perform better than the 3 category scale. Comparisons of MAX values across groups suggested that the 5 category group was marginally, significantly higher than the 3 category group ( $t(56) = 1.66; p < .10$ ), the 7 category group was not different from the 5 category group, and 7 category group was directionally higher than the 3 category group. Comparisons of FUZZY values across groups suggested that the 5 category group was directionally higher than the 3 category group, the 7 category group was not different from the 5 category group, and 7 category group was directionally higher than the 3 category group. These results could be a function of both the number of response categories and the nature of descriptors used to label these categories. The results suggest that both the 5 and 7 category scales may be more appropriate than the 3 category scale.

For the item on frequency of visits to the theater, the values of MAX (0.78, 0.66, and 0.79, respectively for 3, 5, and 7 categories) and FUZZY (0.62, 0.43, and 0.64, respectively for 3, 5, and 7 categories) suggest that both the 3 and 7 category scale perform better than the 5 category scale. Comparisons of MAX values across groups suggested that the 3 category group was significantly higher than the 5 category group ( $t(56) = 1.72; p < .05$ ), the 7 category group was significantly higher than the 5 category group ( $t(57) = 1.91; p < .05$ ), and 7 category group was not different from the 3 category group. Comparisons of FUZZY values across groups suggested that the 3 category group was significantly higher than the 5 category group ( $t(56) = 2.00; p < .05$ ), 7 category group was significantly higher than the 5 category group ( $t(57) = 2.14; p < .05$ ), and 7 category group was not different from the 3 category group. These results suggest that the 5 category scale was the most appropriate of the three options.

The use of more than one response category to indicate a response may be the result of both the mismatch in the number of response categories and category descriptors discussed earlier and the result of aggregating across situations to provide a more complete response. For example if a respondent usually goes to the theater once a month but sometime goes twice a month, membership values in these two response categories would be captured by the scale proposed here. However, the traditional scale would assume that a single response category completely captures the response. This idea of aggregating across time may be of particular relevance for behavioral frequency scales

which involve estimates of frequencies.

-----  
 Insert Table 3 about here  
 -----

### Discussion of Results

Several interesting findings emerge from the study reported here. With respect to behavioral frequency scales, the indicators used here provide a basis for choosing the most appropriate scale. These results could be a function of several factors such as the number of categories used as well as the specific category descriptors used here. For the response-centered scales, it appears that the 7 category scale may be the most appropriate of the three options considered based on differences across scales on the indicators. Since all scales were end-anchored identically, these results can be attributed to the number of response categories in each scale. For the stimulus-centered scale, an argument could be made that the 7 category scale was the most appropriate based on the indicators. However, a significant difference between the 7 category scale and the 5 category scale was not obtained for any of the indicators. Again, the results obtained here can be attributed to the number of response categories in each scale.

-----  
 Insert Table 3 about here  
 -----

## GENERAL DISCUSSION

This paper conceptualized response categories as fuzzy sets to address an important issue in scale development, the optimal number of response categories to use in a scale. Other applications such as the assessment of category descriptors as well as the collection of information on spread or range inherent in some responses were also discussed. A new type of scale was used here which allows for the choice of one or more response categories with the assignment of membership values to each response category. A study is reported where responses to stimulus-centered, response-centered, and behavioral frequency items were collected using this new scale while manipulating the number of response categories across groups of subjects. Using the norm that perfect membership in a single category represents an ideal scenario, several indicators of the appropriateness of scales

were used here. The results are interpreted in terms of recommendations for the choice of the optimal number of response categories.

Information about the membership of responses in more than one response category cannot be inferred from existing measurement procedures which relate to single category responses.<sup>4</sup> Several alternate approaches may be adopted in order to attempt to capture responses more completely. One approach is to develop empirical procedures which allow responses that can belong to multiple categories with varying degrees of membership and incorporate such information into subsequent analyses incorporate it into estimates of reliability and validity. However, such information comes at a cost in terms of the amount of data that needs to be collected and analyzed. A second approach is to assess responses using such scales at the measure development stage in order to make a choice of the most appropriate scale in terms of characteristics such as the number of response categories and category descriptors. Such an assessment could provide a basis for the choice of appropriate scales for the purpose at hand. Several important insights into scale response can be gained by conceptualizing response categories as fuzzy sets and broadening the hitherto narrow perspective that scale response involves the choice of a single response category.



## References

- Cox, Eli (1980), "The Optimal Number of Response Alternatives in a Scale: A Review," Journal of Marketing Research, 17, 407-422.
- Kempton, Willett (1978), "Category Grading and Taxonomic Relations: A Mug is a Sort of a Cup," American Ethnologist, 5, 44-65.
- Lakoff, George (1973), "Hedges; A Study in Meaning Criteria and the Logic of Fuzzy Concepts," Journal of Philosophic Logic, 2, 458-508.
- McCloskey, Michael E., and Sam Glucksberg (1978), "Natural Categories: Well Defined or Fuzzy Sets," Memory and Cognition, 6(4), 462-472.
- Perri, Matthew, and Alan P. Wolfgang (1988), "A Modified Measure of Need for Cognition," Psychological Reports, 62, 955-957.
- Rosch, Eleanor (1973), "On the Internal Structure of Perceptual and Semantic Categories," in Cognitive Development and the Acquisition of Language, ed. T.M. Moore, NY: Academic Press.
- Smithson, Michael (1982), "Models for Fuzzy Nominal Data," Theory and Decision, 14, 51-74.
- Zadeh, L.A. (1976), "A Fuzzy-algorithmic Approach to the Definition of Complex or Imprecise Concepts," International Journal of Man-Machine Studies, 8, 249-291.

## Footnotes

<sup>1</sup> The terms "precise" and "fine-grained" refer to how finely distinguished the values on a continuum are from other possible values. A scale that is sensitive to 1 cm is more fine-grained than a scale that is sensitive to 1 inch, since a 1 cm interval is a finer increment than a 1 inch interval. Restated in terms of the number of response categories used to describe a continuum, if relatively few categories are used (such as the use of 'high', 'medium', and 'low' to describe gas mileage among automobiles), these categories are referred to as being coarse-grained or imprecise and vice versa. These terms are used in a relative sense and do not convey any absolute level of 'grainedness'.

<sup>2</sup> In fact, this process of choosing a point on a relatively fine-grained scale to represent relatively imprecise responses may result in a greater loss of information than the earlier case. Consider a case where a ten point scale is used to measure a five point continuum in memory and the exact reverse (Fig. 1). Since the stimulus value in memory in the latter case is more imprecise, the response generated onto a more fine-grained scale is likely to have a 'wider spread' (or positive membership values with more scale points) than in the former case. However, in the case of the reverse scenario, some responses may be completely captured by a single response category. Therefore, while the fine-grained scale with a single point response may provide an illusion of precision, it may result in greater loss of information of this nature than a coarse-grained scale.

<sup>3</sup> Correlations between MAX and FUZZY for the stimulus-centered items, the response-centered items, and the behavioral frequency items for the 3 category, 5 category, and 7 category groups respectively were 0.76 ( $p < .01$ ), 0.69 ( $p < .01$ ), 0.79 ( $p < .01$ ), 0.75 ( $p < .01$ ), 0.63 ( $p < .01$ ), 0.62 ( $p < .01$ ), 0.84 ( $p < .01$ ), 0.61 ( $p < .01$ ), and 0.74 ( $p < .01$ ).

<sup>4</sup> Measures of reliability represent the primary means of assessing information gained by increasing the number of response categories in traditional measurement. While estimates of reliability are computed from data on single category responses, it could be argued that individual level fuzziness is captured by between variance across individuals at least in the case of stimulus-centered scales (for response-centered scales such a variance would represent trait variance).

Consider a case where a response has equal degrees of membership with three categories. Hence, if a response involves differing memberships in more than one response category, and the choice of any single category from the scale is assumed to be random, it could be argued that an equal number of individuals will chose each of these points. Therefore, the variance in response to that item across individuals is increased. However, such variance confounds individual differences in response with intra-individual *spread* in response. Further, it should be noted that such an argument has merit only with comparable degrees of membership in more than one response category.

## APPENDIX

Consider your response to the following question. "How often do you visit the mall ?"

Once a year	Twice a month	Once a month	Twice a week	Once a week
_____	_____	_____	_____	_____

Now, in the usual kind of scale, you would select anyone of these responses. However, sometimes your response may not be any one of these points but somewhere in between. In other words, none of the alternatives given to you above may capture your response. Take the example when you think your response is somewhere in between twice a month and once a week. You would like to indicate this by checking both alternatives. That is what is possible in this scale in the following way.

If you think that your response is "Completely described by an alternative", you can check that alternative and write down the value "1.0" beneath as shown below. Say, your response is completely described by "Once a week", you will indicate it as shown below.

Once a year	Twice a year	Once a month	Twice a month	Once a week
0.2	0.2	0.2	0.2	0.2
0.4	0.4	0.4	0.4	0.4
0.6	0.6	0.6	0.6	0.6
0.8	0.8	0.8	0.8	0.8
1.0	1.0	1.0	1.0	1.0
_____	_____	_____	_____	_____

If you think that your response is "**mostly described by an alternative**", you can circle the value "**0.80**" above that alternative.

If you think that your response is "**sort of described by an alternative**", you can circle the value "**0.60**" above that alternative.

If you think that your response is "**not too well described by an alternative**", you can circle the value "**0.40**" above that alternative.

If you think that your response is "**not really described by an alternative**", you can circle the value "**0.20**" above that alternative.

If you think that your response is "**not at all described by an alternative**", you do **not** have to circle any value for that alternative (it is equivalent to a value of "0.00").



Table 1RESULTS FOR STIMULUS-CENTERED ITEMS

Item	MAX VALUES			FUZZY VALUES		
	NO. OF RESPONSE CATEGORIES					
	3	5	7	3	5	7
1	0.70	0.72	0.83	0.62	0.65	0.73
2	0.74	0.79	0.75	0.66	0.72	0.61
3	0.66	0.79	0.77	0.57	0.70	0.68
4	0.62	0.66	0.74	0.57	0.55	0.68
5	0.62	0.74	0.69	0.56	0.63	0.63
6	0.66	0.77	0.73	0.59	0.62	0.62
7	0.62	0.60	0.71	0.57	0.57	0.64
8	0.54	0.63	0.77	0.46	0.50	0.68
9	0.74	0.70	0.79	0.67	0.61	0.63
10	0.73	0.72	0.78	0.69	0.62	0.71
11	0.70	0.73	0.79	0.68	0.65	0.69
12	0.65	0.72	0.76	0.61	0.59	0.67
MEAN	0.67	0.72	0.76	0.60	0.62	0.67

Table 2RESULTS FOR RESPONSE-CENTERED ITEMS

Item	MAX VALUES			FUZZY VALUES		
	NO. OF RESPONSE CATEGORIES					
	3	5	7	3	5	7
1	0.68	0.74	0.85	0.53	0.61	0.67
2	0.56	0.70	0.79	0.46	0.59	0.69
3	0.69	0.76	0.78	0.59	0.63	0.70
4	0.66	0.74	0.83	0.55	0.60	0.70
5	0.72	0.72	0.79	0.66	0.59	0.70
6	0.70	0.71	0.83	0.61	0.59	0.71
7	0.67	0.78	0.82	0.59	0.70	0.69
8	0.68	0.75	0.82	0.58	0.66	0.68
9	0.63	0.74	0.82	0.52	0.63	0.65
10	0.67	0.73	0.79	0.61	0.63	0.64
11	0.63	0.70	0.81	0.57	0.61	0.69
12	0.68	0.74	0.85	0.60	0.69	0.76
13	0.57	0.77	0.81	0.48	0.66	0.74
14	0.61	0.72	0.77	0.51	0.59	0.65
15	0.66	0.70	0.82	0.60	0.61	0.63
16	0.68	0.72	0.85	0.62	0.59	0.69
MEAN	0.65	0.73	0.82	0.57	0.62	0.69

Table 3RESULTS FOR BEHAVIORAL FREQUENCY ITEMS

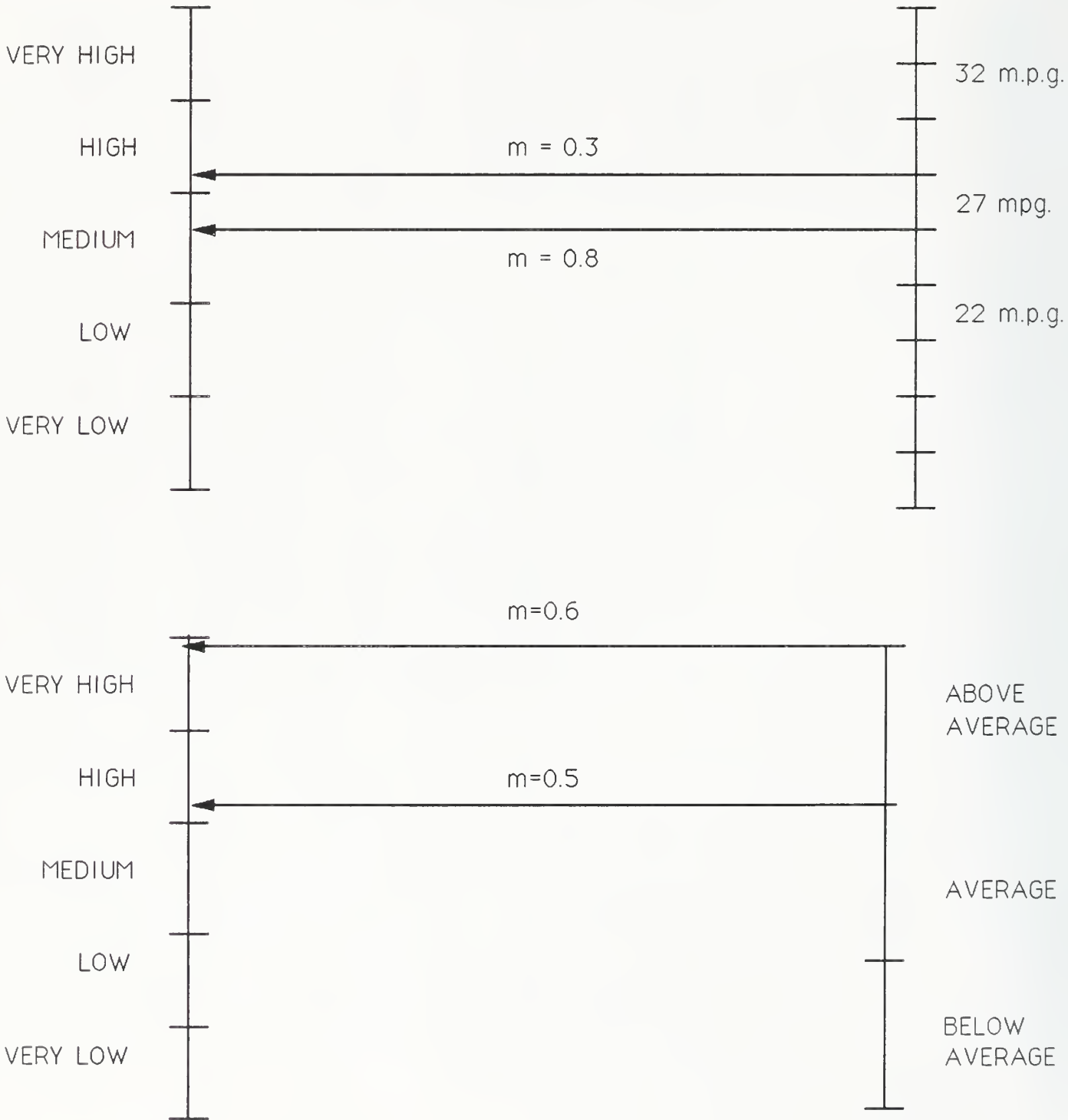
Item	MAX VALUES			FUZZY VALUES		
	NO. OF RESPONSE CATEGORIES					
	3	5	7	3	5	7
1	0.70	0.82	0.79	0.60	0.69	0.71
2	0.78	0.66	0.79	0.62	0.43	0.64

FIGURE 1

MISMATCH BETWEEN THE NUMBER OF RESPONSE CATEGORIES AND SCALE RESPONSES

MEASUREMENT SCALE

SCALE RESPONSE















HECKMAN  
BINDERY INC.



**JUN 95**

Bound-To-Please® N. MANCHESTER,  
INDIANA 46962



UNIVERSITY OF ILLINOIS-URBANA



3 0112 004996879