Does a Single Response Category in a Scale Completely Capture a Response?

Madhubalan Viswanathan
University of Illinois at Urbana-Champaign

Mark Bergen and Shantanu Dutta
University of Chicago

Terry Childers
University of Minnesota

ABSTRACT

The objective of this article is to suggest the application of fuzzy set theory to issues in scale development in marketing by conceptualizing response categories of a scale as fuzzy sets. Respondents' natural responses to scales are argued to have degrees of membership in more than one response category when a response scale does not *match* natural responses. In contrast to traditional scales, which only allow the choice of a single category as a response, a new scale, which allows responses with degrees of membership in more than one response category, is developed. This scale is used in an expository study conducted to demonstrate implications of the conceptualization for issues in scale development. Specifically, norms are suggested for use in evaluating traditional scales during scale development in terms of the optimal number of response categories and appropriate category descriptors. Implications of this methodology for marketers are discussed relating to the use of the proposed methodology in (a) the development of response scales, (b) the development of questions/items in questionnaire design, and (c) the measurement of phenomena that are inherently imprecise.

Psychology & Marketing
Vol. 13(5):457-479 (August 1996)
CCC 0742-6046/96/130457-23

The use of scales to measure responses is pervasive in marketing. Typically, these scales provide several response categories and respondents are required to choose any one that captures their response. For example, respondents may be asked to indicate the number of hours a day they watch television, on a scale with response categories such as "Up to 1 hour," "1–2 hours," "2–4 hours," and "More than 4 hours." Scales such as this, which are traditionally used, capture the extent of television viewing on the basis of the single response category indicated by the respondent. However, a respondent who watches television for 1–3 hours daily may have to choose between "1–2 hours" and "2–4 hours." In other words, a single response category may not completely capture a response if the sets of response categories provided do not match the natural responses of respondents. Though the example above related to the category labels in a scale, such problems may arise because of several characteristics of the response scales, such as the number of response categories. For example, response categories of scales used to measure attitudes and purchase intentions may not completely capture responses. Traditional scales, by requiring the choice of a single response category as a response, cannot provide information on the extent to which a response category captures respondents' natural responses.

This article suggests the application of fuzzy set theory (cf. Kaufman, 1975; Zadeh, 1976) to scale development by conceptualizing response categories of a scale as fuzzy sets. A fuzzy set, as opposed to a crisp set, is one whose members can possess varying degrees of membership rather than either belonging or not belonging to the set (Zadeh, 1976). As suggested by the example above, respondents' natural responses to scale items (i.e., responses to an item that respondents would provide in the absence of restrictions imposed by a response scale, in terms of, say, the number of response categories or the specific set of category descriptors used) may have degrees of membership in more than one response category when a response scale does not *match* these natural responses. Mismatches are argued to occur because of differences between the respondents' stored natural responses on a continuum and response categories of a scale used to represent that continuum. Because traditional scales require the choice of a single response category as a response, information on membership of responses in more than one category cannot be elicited.

In contrast to traditional scales, a new scale is proposed that allows for responses that indicate degrees of membership in more than one response category. This new scale is argued to be useful during the development of traditional scales to provide diagnostic information about issues such as the optimal number of response categories to use in a scale in order to minimize loss of information due to mismatches between natural responses and response scales. Based on the extent to which respondents use more than one response category to indicate

their response to an item, norms are suggested for evaluating and developing appropriate traditional scales in marketing. A study was conducted to demonstrate the use of this fuzzy set methodology in scale development. With the use of the new scale described above, data were collected for stimulus-centered, subject-centered, and completely described scale items. Different numbers of response categories are used across groups of subjects. The study emphasizes specific issues in scale development for illustrative purposes. However, broader implications of the conceptualization and methodology for marketers are discussed subsequently and relate to the use of the proposed methodology in (a) the development of response scales, (b) the development of questions/items in questionnaire design, and (c) the measurement of phenomena that are inherently imprecise.

RESPONSE CATEGORIES OF A SCALE AS FUZZY SETS

This section briefly describes the application of fuzzy sets to scale development by conceptualizing response categories as fuzzy sets. Implications of this conceptualization for issues in scale development are then discussed. Scenarios where a response may have degrees of membership in more than one response category are explained in terms of mismatches between natural responses and response scales. The implications of such mismatches for issues in scale development such as the optimal number of response categories and the type of category descriptors are discussed. A new scale that can capture degrees of membership in multiple response categories is also discussed.

Application of Fuzzy Sets to Response Categories

Zadeh (1976) suggested the notion of a fuzzy set, a concept that has been applied in several areas, such as in assessing membership of instances in natural categories (McCloskey & Glucksberg, 1978) and in coding qualitative data (Smithson, 1982a). Zadeh's explanation of the nature of fuzzy sets can be understood with the use of a frequently encountered problem in marketing, determining the optimal number of categories to use in a scale. Researchers often face a trade-off between the finer discriminating power achieved by increasing the number of scale response categories versus the additional burden and fatigue that is placed upon the respondent and the resulting quality of response (cf. Cox, 1980). However, complicating the need for decreasing the number of response categories is the need for respondents to be able to make adequate discriminating judgments, because providing too few categories may result in response ambiguity. This would impact the psychometric qualities of the measurement scale through a reduction in reliability and validity.

Consider a typical rating scale used to measure ratings of gas mileage of automobiles with the use of three response categories, "high," "medium," and "low." Say respondents are aware of gas mileage of automobiles to the nearest mile(s) per gallon (mpg). Considering their responses with respect to the response category "high," many respondents may rate 24 mpg as definitely not being high mileage and 40 mpg as definitely being high mileage. However, a certain number of mpg above 24 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 respondents are assumed to set an arbitrary criterion such that any mileage that is 1 mpg greater than 24 is considered high, then the distinction between high and not high (i.e., medium or low) reduces to being equivalent to 1 mpg, raising the issue as to where a magnitude such as 24.5 mpg would belong. If large intervals such as 5 mpg are used to set a criterion, then the intermediate range of mileages (from say, 24 to 29 mpg) is undefined. The use of an arbitrary criterion to define an inherently imprecise category leads to minute distinctions between high and not high. Zadeh (1976) attempted to resolve this paradox by using the notion of fuzzy sets. If Zadeh's explanation is applied to the present example, terms such as *high* are vague or imprecise and there is a gradual transition from mpg that are not high to mpg that are high. A category such as "high" is called a fuzzy set (as opposed to a crisp set) because it eliminates the sharp distinction between members and nonmembers and allows for *gradations* 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" (as well as in the categories "medium" and "low"), with higher values representing greater degrees of membership.

If this reasoning is extended to a set of response categories in a scale, a categorical scale that is typically used in research in marketing involves the use of a group of response categories or fuzzy sets to capture responses along some continuum. 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 scales 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 of a response in more than one response category. For example, 32 mpg may be considered as belonging to the category "high" with a membership of 1.0, whereas 28 mpg may be considered as belonging to the category "high" with a membership of 0.6 and the category "medium" with a membership of 0.3. However, the use of a traditional scale would lead to both 32 and 28 mpg being classified as high. (It should be noted that such a prob-

lem cannot be resolved by using category descriptors such as "high [above 24 mpg]" because it would not capture the subjective impression of highness as perceived by a respondent.) The key point here is that response categories may be inherently fuzzy or imprecise in nature and that responses may be partial members in one or more categories. Therefore, the argument advanced is that response categories are similar to natural categories in terms of allowing graded membership (cf. Rosch, 1973). Gradedness in natural categories has been argued to occur because of various combinations of featural and dimensional values leading to a continuum of membership in a category. We argue that graded membership of responses in response categories occurs because of *mismatches*, defined as the difference between the respondents' stored natural responses on a continuum and response categories used to represent that continuum. It should be noted that responses with degrees of membership in multiple categories due to mismatches do not represent *response error* but the *spread* or *range* of a response to an item.

Implications for Issues in Scale Development

Viewing response categories as fuzzy sets allows insights about issues in scale development such as the optimal number of response categories to utilize in a scale. *Mismatches* between natural responses and response scales will be argued to lead to responses that have degrees of membership in more than one response category. We suggest that such responses may be truncated by traditional scales which require single category responses. Therefore, one approach is to minimize the loss of information due to such responses by maximizing the match between traditional response scales and natural responses during scale development. This diagnostic information can then be used to develop traditional scales where the response categories have been selected to minimize the degree of mismatch. This fuzzy set methodology will be demonstrated with the use of two issues in scale development: the number of response categories and the set of category descriptors to use in scales.

Optimal Number of Response Categories to Use in a Scale. Several researchers in marketing and psychology have studied the problem of the number of categories to use in a scale (Cox, 1980; Komorita & Graham, 1965). Suggestions made by researchers range from the use of 2–25 alternatives (Cox, 1980). Although the use of seven response categories is often cited as being ideal for measurement scales because human ability to discriminate 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, 1956) and may not be generalizable to other contexts.

Therefore, 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. In this regard, Cox (1980) suggests that there is an immediate need to develop methods at the pretesting or scale development stage to evaluate the nature of information being collected when using different numbers of response categories.

Scenarios where responses may have degrees of membership in multiple response categories are discussed here with the use of the concept of a *mismatch* between natural responses and response categories in terms of precision or fine grainedness. The terms *precise* and *fine grained* refer to how finely distinguished the values on a continuum are from other possible values. A scale measuring gas mileage of automobiles that is sensitive to 1 mpg is more fine grained than a scale that is sensitive to 5 mpg, because a 1-mpg interval is a finer increment than a 5-mpg 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. Two possible scenarios will be considered wherein natural responses are more fine grained and less fine grained, respectively, than response scales.

Consider a scenario where respondents' natural responses are more fine grained than the response scales used to measure them [see Figure 1(a), where natural responses involve five response categories, whereas the response scale allows only three response categories]. Because relatively fine-grained responses have to be reduced to fit a set of relatively coarse-grained response categories, no single response category may completely capture a response. Rather, the response may have varying degrees of membership in more than one response category. For example, in Figure 1(a), the response "high" does not fit completely into any response category, but overlaps with two categories to different degrees. Such overlap occurs because of a mismatch between the categories used in the response scale and respondents' natural responses, thereby leading to the possibility of graded membership of responses in one or more of these categories.

A similar problem exists if response categories are more fine grained than the respondent's natural responses [see Figure 1(b)]. A relatively coarse-grained response such as "above average" mileage overlaps with two categories on the response scale (i.e., "high" and "very high"), leading to the possibility of membership in each of these two categories. The problem here is the reverse, to match relatively coarse-grained responses to a more fine-grained scale. Whereas more than one response category may be chosen for any particular response, traditional scales restrict the choice of responses to a single category.

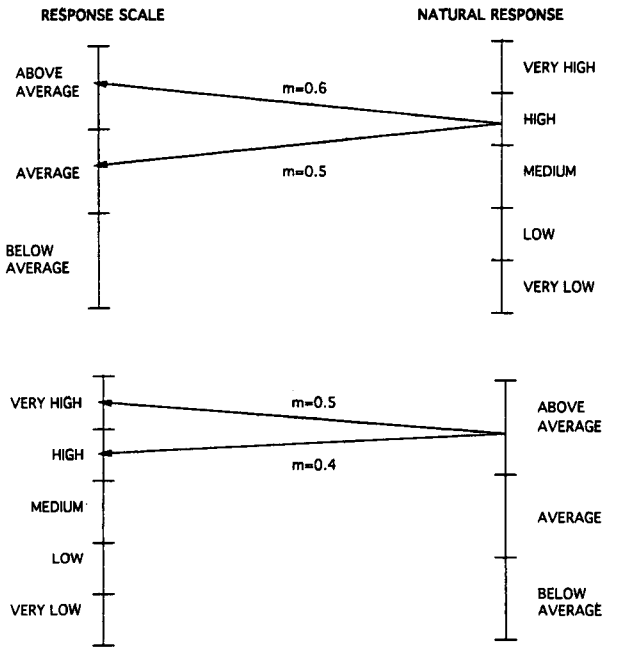

Figure 1 Types of mismatches between natural responses and response scales. (a) m = 0.6, m = 0.5. (b) m = 0.5, m = 0.4.

Hence, as long as there is a mismatch in terms of the number of categories available between the respondent's natural categories of stored knowledge and the responses to a scale, there is loss of information due to the traditional requirement of a single-category response. In this scenario, the loss of information is due to a relatively coarse-grained response being captured by a relatively precise response category. Note that the use of precise response categories does not solve the problem of multiple category membership of responses because the responses do not match the response scale in terms of precision. In fact, this process of choosing a category on a relatively fine-grained scale to represent relatively coarse-grained responses may result in a greater loss of information than the earlier case. Consider a case where a five-category scale is used to measure a set of natural responses consisting of three categories [Figure 1(b)] and the exact reverse [Figure 1(a)]. Because respondents' natural responses in the former case are more coarse grained, the response generated onto a more fine-grained scale is likely to have a wider spread (or positive membership values with more response categories) than in the latter case. However, in the case of the latter [Figure 1(a)], some responses may be completely captured by a single response category (for example, the responses "very high" and "very low"). Therefore, though a fine-grained scale may appear to capture fine-grained responses, it may result in greater loss of information through spread in response as compared to a coarse-grained scale.

The discussion to this point assumes that respondents' natural responses may vary in terms of level of precision or fine grainedness. Such a premise is supported by past research, which suggests that the degree of precision or fine grainedness with which respondents store information on an attribute continuum may vary as a function of several factors. Johnson and Fornell (1987) suggest that relatively concrete attributes may be stored as features (which have only two categories, the presence or absence of a feature), whereas more abstract attributes may be stored as dimensions (which may have more than two categories) in memory. Alba and Hutchinson (1987) suggest that experts can make more memory-based discriminations along dimensions than novices. Additionally, Park and Lessig (1981) report directional support for the prediction that moderate and high product familiarity leads to the use of narrower categories on an attribute (which was argued to suggest more categories) when compared to low product familiarity.

Given the nature of responses that may arise due to mismatches, responses collected on a scale that allows responses with varying degrees of membership in multiple response categories could provide important diagnostic information during scale development about the optimal number of response categories to use in that scale. Such information could be obtained by varying the number of categories on such a scale during development and comparing the extent to which more than one response category is utilized by respondents for a set of items. Ideally, to the extent that respondents tend to use a single response category with perfect membership to characterize their response, the number of response categories used in a scale can be considered as being appropriate in terms of avoiding loss of information due to membership of responses in more than one category. As responses approach the ideal described above, the number of response categories used could be argued to be more and more appropriate (in terms of less loss of information due to membership of responses in multiple categories), thereby providing a basis to choose between scales with different numbers of response categories. Hence, the tradeoffs between using scales that are too fine grained versus too coarse grained in determining the optimal number of response categories is approached in terms of the degree to which responses may have memberships in more than one response category.

Category Descriptors to Use in a Scale. Another issue in scale development that is addressed is a mismatch between natural responses and response categories in terms of descriptors used to label response categories. Consider an item with a scale whose response categories are completely described (e.g., for an item on the hours of daily television viewing, a set of labels such as "Up to 1 hour," "1–2 hours," and "More than 2 hours"). To the extent that the set of labels does not

match the natural responses of respondents, membership of responses in more than one category may occur. A respondent who watches television for 1½–2½ hours daily may have to choose both "1–2 hours" and "More than 2 hours," with some degree of membership in each. Such responses, which arise due to a mismatch between the set of descriptors used in a scale and the responses provided, could be examined during scale development. This could be done here by varying the descriptors on a scale and studying the extent to which more than one response category is utilized by respondents.

METHOD

For expository purposes, a study was conducted that illustrates the manner in which fuzzy set theory can be applied within the area of scale development. Not all aspects of scale development nor a complete assessment of the issues we investigated are addressed in this study. Rather, the objective is to demonstrate the nature of the proposed methodology.

Development of a Scale Based on the Fuzzy Set Approach

A scale that assesses the extent to which a response was captured by that category by allowing responses that can have degrees of membership in more than one response category was used here. This scale was derived from past research (Smithson, 1982a) which used a fuzzy set theoretic framework to develop techniques for coding qualitative data. In coding tasks, observations are usually classified into predetermined categories. However, in open-ended interviews or in field studies involving observations of behavior, observations may not precisely fit a single category (Smithson, 1982a). Researchers have suggested phrases that can be used to collect information about degrees of membership of items in categories (cf. Kempton, 1984; Lakoff, 1973). Smithson (1982a) used a range of phrases to indicate degrees of membership of items in categories presented by Kempton (1978), and suggested the assignment of membership values to items to indicate their memberships to various categories. The suggested phrases and corresponding membership values were 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 by 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, 1982a). By the application of this scheme to response categories of a scale, a new scale was used in this study that allowed respondents to assign degrees of membership to each response category to indicate the extent

to which a response was captured by that category. The phrases suggested by Smithson (1982a) as well as the corresponding levels of membership were used with the replacement of the phrase *coding scheme* with the word *alternatives* (see the Appendix for a sample scale and instructions).

Overview and Procedures

The approach employed in this study was to use the scale described above to collect responses for different items from groups of subjects assigned to conditions with varying numbers of response categories. Three groups of 30 students each at a midwestern university were assigned to conditions where three, five, and seven response categories, respectively, were used for scales. The aim here was to assess several different scales that are commonly used in marketing, that is, scales with three, five, and seven response categories. Data were collected on three types of items, namely, stimulus-centered items, subject-centered items, and an item with a completely described scale (referred to as a completely described scale item), with the use of a questionnaire format. Cox (1980) points out that research in marketing on the optimal number of categories to use in a scale has involved stimulus-centered and subject-centered items, with the stimulus-centered item being one where the variation in responses is attributed to differences between stimuli on an attribute and the subject-centered item being one where variation in responses is attributed to differences between respondents. Having completely described scale items allows a demonstration of the application of the fuzzy set approach to determine the category descriptors to use in a traditional scale. Responses to stimulus-centered items involved rating how much respondents liked a set of 12 soft drinks on scales anchored by "Very Bad" and "Very Good." Responses to subject-centered items involved the use of a 16-item version of the Need for Cognition scale (Perri & Wolfgang, 1988, with items such as "I am an intellectual") with scales that were anchored by "Strongly Disagree" and "Strongly Agree." The completely described scale item was on hours of daily television viewing with completely described response categories (anchored for three, five, and seven category scales, respectively, as follows: "Up to 1 hour," "1–2 hours," and "More than 2 hours;" "Up to ½ hour," "½–1 hour," "1–1½ hours," "1½–2 hours," and "More than 2 hours;" "Up to ½ hour," "½–1 hour," "1–1½ hours," "1½–2 hours," "2–2½ hours," "2½–3 hours," and "More than 3 hours").

Subjects were provided with detailed instructions regarding the scales and completed several sample trials. The instructions followed Smithson (1982a) 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. A variation of the scale that 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. Nonresponse to a response category indicated a membership value of 0.0. Subjects completed the stimulus-centered, subject-centered, and completely described scale items in that order, followed by items about their reactions to the use of the new fuzzy-set-based scale.

Mean ratings across all 90 subjects for the items tapping subjects' reactions in completing the new scale appeared to be satisfactory and were as follows: motivation to complete scales (10-point scale anchored by "Not at all motivated" and "Very motivated"; 6.3/10), knowledge level to complete scales (10-point scale anchored by "Very low," and "Very high"; 7.5/10), familiarity with completing scales (10-point scale anchored by "Very low" and "Very high"; 5.8/10), adherence to instructions (10-point scale anchored by "To a large extent" and "Not at all"; 5.6/10 after reverse scoring), confidence in responses provided (10-point scale anchored by "Very low" and "Very high"; 7.0/10), satisfaction with accuracy of responses (10-point scale anchored by "Very dissatisfied" and "Very satisfied"; 6.9/10), certainty in responses (10-point scale anchored by "Not at all certain" and "Very certain"; 6.8/10), sureness in responses (10-point scale anchored by "Not at all sure" and "Very sure"; 6.9/10), and ease of completing scales (10-point scale anchored by "Very difficult" and "Very easy"; 6.3/10). These results suggest that the new scale was completed with moderate levels of motivation, familiarity, adherence to instructions, and ease, and moderately high levels of knowledge, as well as confidence, perceived accuracy, certainty, and sureness in responses.

Data Analysis and Results

Preliminary data analysis of responses to different types of items were performed to assess the degree to which respondents used more than one response category to indicate their responses. Because the premise of this research was that responses may not be fully captured by a single response category, the aim here was to empirically assess the extent to which respondents used multiple response categories to indicate their responses. The mean number of response categories of a scale to which respondents assigned nonzero membership values were computed for each set of items. For the condition with the 3-point scale, the mean number of response categories used for the stimulus-centered, subject-centered, and completely described items were 1.16, 1.25, and 1.28, respectively, all significantly greater than 1, which would have been the default mean if only a single response category had been used. For the condition with the 5-point scale, the mean number of response categories used for the stimulus-centered,

subject-centered, and completely described items were 1.19, 1.25, and 1.28, respectively, all significantly greater than 1. For the condition with the 7-point scale, the mean number of response categories used for the stimulus-centered, subject-centered, and completely described items were 1.19, 1.30, and 1.13, respectively, all significantly greater than 1. These results suggest that respondents used the new scale to indicate responses that had membership in more than one response category, providing empirical evidence for the premise of this research that scale responses may have membership in multiple response categories of traditional scales.

Indicators for Evaluating Scales. Responses belonging to a single category with a membership of 1.0 represent an ideal scenario, because traditional scales allow the choice of only a single-category response. Norms were used to capture the degree to which scale responses were close to this ideal. Indicators of distance from this ideal measured the extent to which a *single* category completely captured the response to an item. The primary indicator used here, referred to as a fuzzy index (FUZZY), was the distance between a set of membership values provided in a response and the nearest ideal set, which was the set with a membership value of 1 in one response category and 0 in all other categories (i.e., similar to a response on a traditional scale). This indicator was adapted from an index of fuzziness in fuzzy set theory that measures the distance between a fuzzy set and the nearest crisp set (Kaufman, 1975). An example of the computation of FUZZY is shown below with the use of a response to a 3-point scale with membership values of m_1, m_2, and m_3 in the three response categories (say, 0.8, 0.2, and 0.0, respectively, such that m_1 was the highest membership value, also referred to here as MAX). The computational formula was $\text{ABS}(1 - \text{MAX}) + \text{ABS}(0 - m_2) + \text{ABS}(0 - m_3)$, with ABS referring to absolute values, because the nearest ideal set would have membership values of 1, 0, and 0, for m_1, m_2, and m_3, respectively. It should be noted that the ideal set corresponds to the response that would have been obtained with a traditional scale. If one response category completely captured a response, then this index would be 0, with higher values suggesting greater distance from the ideal set.

A secondary indicator used here was the maximum membership value (MAX) that was assigned to any of the response categories of a scale. This indicator was based on the rationale that higher membership values in any one response category, being closer to the ideal membership value of 1.0, are indicative of a more appropriate scale. Though this indicator does not completely capture the notion of distance of a response from the ideal set, it was explored here based on the premise that a high membership in one category should lead to a low FUZZY value. Correlations between FUZZY and MAX, for the three-category, five-category, and seven-category groups, respectively,

were −0.76, −0.69, and −0.85 for the stimulus-centered items, −0.75, −0.63, and −0.69 for the subject-centered items, and −0.84, −0.61, and −0.82 for the completely described scale item. All these correlations were significant at the 0.01 level, suggesting convergence between these two indicators.

Results for Stimulus-Centered Items. The mean fuzzy values (FUZZY) and maximum values (MAX) were computed for each item for each condition with respect to the number of response categories in a scale. Further, the means of these indicators across the set of 12 items were also computed (see Table 1). Therefore, response scales were evaluated in terms of natural responses to a *set* of items, because the notion of a mismatch was developed in terms of a continuum of responses rather than an individual response. One-way ANOVAs were run on FUZZY and MAX and results of overall F tests are presented below in parentheses, along with the means. As evident from the overall mean and the means for several items, the values of FUZZY decreased [0.40, 0.38, and 0.33, respectively, for three, five, and seven categories; $F(2,85) = 0.92$], and the values of MAX increased [0.67, 0.72, and 0.76, respectively, for three, five, and seven categories; $F(2,85) = 2.99$; $p < .06$], with an increase in the number of response categories. No significant differences were obtained for comparisons of FUZZY values across groups, with the seven-category group being directionally lower than the other two groups. Comparisons of MAX values across groups suggested that the five-category group was marginally higher than the three-category group [$t(56) = 1.32$; $p < .10$], with the seven-category group being directionally higher than the five-category group [$t(57) = 1.18$; $p > .10$], and significantly higher than the three-category group [$t(57) = 2.35$; $p < .05$]. Because all scales had identical end anchors, these results can be attributed to the number of response categories in each scale. It appears based on these results that seven response categories may be the most appropriate among the three options considered. If one speculates on the pattern of results in terms of the decrease in FUZZY and increase in MAX with an increase in the number of response categories, it appears that the type of mismatch occurring here is due to natural responses being more precise than the response scale. Therefore, a scale with more than seven categories may perform better than any of these three options, perhaps because natural responses to items tapping an overall global judgment of degree of liking may be more precise or discriminating than any of the three options considered. Such a possibility could be investigated by examining responses to scales with more than seven response categories, looking for a U-shaped relationship of FUZZY (or an inverted U-shaped relationship of MAX) with the number of response categories, and identifying the number of response categories that correspond to the lowest FUZZY value.

Table 1. Summary of Results.

	Fuzzy Values			Max Values		
	No. of Response Categories					
Item	3	5	7	3	5	7
	Results for stimulus-centered items					
1	0.38	0.35	0.27	0.70	0.72	0.83
2	0.34	0.28	0.39	0.74	0.79	0.75
3	0.43	0.30	0.32	0.66	0.79	0.77
4	0.43	0.45	0.32	0.62	0.66	0.74
5	0.44	0.37	0.37	0.62	0.74	0.69
6	0.41	0.38	0.38	0.66	0.77	0.73
7	0.43	0.43	0.36	0.62	0.60	0.71
8	0.54	0.50	0.32	0.54	0.63	0.77
9	0.33	0.39	0.37	0.74	0.70	0.79
10	0.31	0.38	0.29	0.73	0.72	0.78
11	0.32	0.35	0.31	0.70	0.73	0.79
12	0.39	0.41	0.33	0.65	0.72	0.76
MEAN	**0.40**	**0.38**	**0.33**	**0.67**	**0.72**	**0.76**
	Results for subject-centered items					
1	0.47	0.39	0.33	0.68	0.74	0.85
2	0.54	0.41	0.31	0.56	0.70	0.79
3	0.41	0.37	0.30	0.69	0.76	0.78
4	0.45	0.40	0.30	0.66	0.74	0.83
5	0.34	0.41	0.30	0.72	0.72	0.79
6	0.39	0.41	0.29	0.70	0.71	0.83
7	0.41	0.30	0.31	0.67	0.78	0.82
8	0.42	0.34	0.32	0.68	0.75	0.82
9	0.48	0.37	0.35	0.63	0.74	0.82
10	0.39	0.37	0.36	0.67	0.73	0.79
11	0.43	0.39	0.31	0.63	0.70	0.81
12	0.40	0.31	0.24	0.68	0.74	0.85
13	0.52	0.34	0.26	0.57	0.77	0.81
14	0.49	0.41	0.35	0.61	0.72	0.77
15	0.40	0.39	0.37	0.66	0.70	0.82
16	0.38	0.41	0.31	0.68	0.72	0.85
MEAN	**0.43**	**0.38**	**0.31**	**0.65**	**0.73**	**0.82**
	Results for completely described scale item					
1	**0.40**	**0.31**	**0.29**	**0.70**	**0.82**	**0.79**

Results for Subject-Centered Items. For subject-centered items, the values of FUZZY decreased [0.43, 0.38, and 0.31, respectively for three, five, and seven categories; $F(2,85) = 3.41$; $p < .05$], and the values of MAX increased [0.65, 0.73, and 0.82, respectively, for three, five, and seven categories; $F(2,85) = 13.33$; $p < .001$], with an increase in the number of response categories (see Table 1). Comparisons of FUZZY

values across groups suggested that the five-category group was marginally lower than the three-category group [$t(56) = 1.54$; $p < .10$], with the seven-category group being directionally lower than the five-category group [$t(57) = 1.24$; $p > .10$], and significantly lower than the three-category group [$t(57) = 2.50$; $p < .01$]. Comparisons of MAX values across groups suggested that the five-category group was significantly higher than the three-category group [$t(56) = 2.60$; $p < .01$], with the seven-category group being significantly higher than both the five-category group [$t(57) = 2.77$; $p < .01$], and the three-category group [$t(57) = 5.48$; $p < .01$]. These results suggest that the seven-category scale may be the most appropriate of the three options considered. Again, the results suggest a type of mismatch where natural responses may be more fine grained than the response scale. Therefore, a scale with more than seven response categories may be more appropriate than a seven-category scale.

Results for the Completely Described Scale Item. For the item on hours of television viewing, the values of FUZZY [0.40, 0.31, and 0.29, respectively for three, five, and seven categories; $F(2,85) = 0.79$] and MAX [0.70, 0.82, and 0.79, respectively for three, five, and seven categories; $F(2,85) = 1.29$] suggest that both the five- and seven-category scale perform better than the three-category scale (see Table 1). Comparisons of FUZZY values across groups suggested that the five-category group and seven-category group were directionally lower than the three-category group, and the seven-category group was approximately equal to the five-category group. Comparisons of MAX values across groups suggested that the five-category group was marginally higher than the three-category group [$t(56) = 1.66$; $p < .10$], with the seven-category group being approximately equal to the five-category group, and directionally higher than the three-category group. It should be noted that these results should be interpreted in light of both the number of response categories *and* the specific category descriptors used, because these scales were completely described with different sets of descriptors. The results suggest that both the five- and seven-category scales with their specific sets of descriptors may be more appropriate than the three-category scale with its specific descriptors.

The results in terms of each of the norms described above also provide estimates of the degree to which respondents used more than one response category for all three types of scales. As shown in Table 1, the means for these norms were sizably different from their respective ideal values. Therefore, similar to the preliminary data analysis, these results provide empirical evidence for the premise of this research that scale responses may have membership in multiple response categories of traditional scales.

GENERAL DISCUSSION

This article suggested the conceptualization of response categories of a scale as fuzzy sets. The proposed conceptualization is a fundamentally different way of viewing scales and their responses. Applications of this conceptualization to address important issues in scale development such as the optimal number of response categories and the set of category descriptors to use in a scale were illustrated here. With the use of this conceptualization, mismatches between natural responses and traditional response scales due to the number of response categories of a scale as well as the category descriptors were argued to lead to responses with degrees of membership in more than one response category. A new type of scale that uses fuzzy set methodology and captures such responses was used in a study to demonstrate applications of the conceptualization. With the use of the norm that perfect membership in a single category represents the ideal (i.e., traditional scales allow the choice of only a single response category), indicators of closeness to this ideal were used to assess scales in terms of the number of response categories and category descriptors. Based upon the approach demonstrated in our study, researchers developing new scales could determine more objectively the optimal number of categories and types of scale descriptors to use in order to reduce loss of information due to mismatches.

Implications for Marketing

This article illustrates the application of the proposed conceptualization to specific issues in scale development. However, there are broader implications of this research for marketing. Among these are applications concerned with addressing problems of question wording and question content in the design of questionnaires, and capturing content domains that are inherently imprecise. Each of these areas of applications is discussed below, followed by a brief discussion of the costs and benefits of this methodology.

Applications in the Development of Response Scales. In terms of application to issues in scale development, as illustrated in detail here, issues such as the optimal number of response categories and optimal category descriptors can be examined during scale development. Both subjective and objective category labels (e.g., scales of income, etc.) can be evaluated through this procedure. The inappropriateness of one or more specific category descriptors in a scale can be diagnosed through this methodology. Although the type of stimulus-centered scale examined here was a liking scale, the proposed methodology could also be applied to attribute rating scales as well as scales measuring purchase and other behavioral intentions. The practical signifi-

cance of using this methodology is in developing scales that validly measure phenomena with the use of response categories that reflect respondents' thought processes. Therefore, scales developed through the benefit of this process may be more predictive of specific criterion variables than other scales. This is because natural responses provided by respondents are more likely to be reflected in their decision making and behavior than responses that are merely the outcome of a measurement procedure imposed on respondents. Viswanathan and Sudman (1995) maintain a similar line of reasoning in a study that uses pretests to assess the number of response categories along certain attributes that are meaningful to respondents. The authors demonstrate that attribute rating scales with a meaningful number of response categories determined on the basis of pretesting (i.e., that reflect respondents' natural responses) perform better than other scales in capturing meaningful differences in overall liking. Similarly, differences on scales that reflect natural responses in measuring overall liking and purchase intentions may be more likely to capture differences that are meaningful to respondents than more traditional scales.

Although the proposed approach was applied to scales that have at least ordinal-level properties, it may be particularly important for nominal scales. Lehmann and Hulbert (1972) contend that there is rouding error when respondents are forced to use a nominal scale to express an attitude that is continuous in nature (such as a 60% Yes and a 40% No on a scale with Yes/No responses). Therefore, in situations where the use of a nominal scale is in question, the fuzzy-set-based scale could be used to collect preliminary data and assess whether the phenomenon in question is amenable to nominal scales or whether additional response categories need to be provided. When using qualitative categories as well, the extent to which a response may be a member in more than one response category needs to be assessed. For example, for a question about race (or nationality or a host of other variables) using conventional response categories, people of a mixed racial background may indicate responses that belong in more than one response category. This methodology could also be used to evaluate the degree to which response alternatives in the form of statements completely capture responses. For example, response alternatives may suffer from wording problems that lead to respondents choosing more than one alternative with nonzero membership and/or that lead to no single alternative completely capturing a response.

Applications in Questionnaire Development. The proposed methodology can also be used to evaluate and modify questions and items in questionnaire design. Such applications would be based on the premise that spread in response occurs because of a lack of certainty in response stemming from problems with question wording and content. Several problems in question wording and question

content can lead to responses with membership in multiple categories in a fuzzy-set-based scale. Complex questions, ambiguous questions, and double-barreled questions are examples of instances where respondents who are uncertain of their response may choose more than one response category. These types of questionnaire design flaws have been difficult to detect through traditional forms of pretesting (Hunt, Sparkman, & Wilcox, 1982) and may benefit from a fuzzy-set-based approach. Similarly, questions that respondents do not have the ability to answer, such as those requiring difficult estimates, may also lead to responses with membership in multiple categories. Wording for an item that is too general and lacks specificity may also lead to responses with membership in multiple categories.

It is noteworthy that, in these cases, the spread in response occurs because of a *lack of certainty* on the part of the respondent rather than mismatch between natural responses and response categories. The notion of mismatch assumes that respondents have a certain response but are unable to express it due to mismatch between their response and the response categories provided by a scale. However, the issues described above regarding problems in questions wording and content relate to situations where a single alternative should suffice to capture a response. In this context, however, problems with the questions lead to lack of certainty in response that is reflected in a greater spread in the response.

Several of the problems discussed above should lead to an element of randomness in the response provided on traditional scales that require the choice of a single response category. This in turn should result in a problem with unreliability. Likewise, with the use of one or two items to measure a construct, traditional reliability procedures are either not possible (with the exception of test-retest) or are rarely carried out on such questions. Typically, problems with question wording during questionnaire design are addressed deductively (i.e., through content validity) without any form of empirical support, or through the use of think-aloud protocol procedures. However, the fuzzy-set-based scale could be used at the pretesting stage to identify these problems empirically. In essence, if spread in response stems from lack of certainty, it represents the randomness that is attached to an individual response. Whereas internal consistency measures randomness across a set of items, and stability measures randomness across time, the randomness captured by a fuzzy set scale provides *a direct measure of unreliability in an individual response to a single item*. Hence, the proposed methodology can be used in questionnaire development in a large number of situations where single-item scales are used to measure a phenomenon at a single point in time. Whereas think-aloud and pilot testing procedures provide valuable qualitative information to guide the design of questionnaires, the fuzzy-set-based

scale offers a way of identifying and quantifying these problems during the development of questionnaires.

Applications in the Measurement of Imprecise Phenomena. Other applications of this methodology include the assessment of responses that may *inherently* involve more than one response category. Such applications would involve the substitution of traditional scales with fuzzy-set-based scales in all phases of data collection. Several examples of inherently imprecise responses illustrate these applications. For an item about the frequency of visits 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 of these categories. Specifically, the response requires some form of aggregation across dimensions such as situations or time. This represents a scenario where the response inherently involves multiple categories, *irrespective of how precise the categories are or how they are labeled.* Similarly, a question about the price of a product may require some aggregation across time, depending on the various promotions that may be offered over a period of time. Traditional scales are not designed to capture such phenomena, but a fuzzy-set-based scale can be readily adapted to such a context. Similarly, a response to a subject-centered item such as "I am an intellectual" may require some form of aggregation across, perhaps, the various roles played by the individual that relate to this item. Such aggregation may be more likely for general rather than specific items because of the need to aggregate across specific situations. Because different respondents may have different imprecise categories as their natural responses, it may not always be viable to use fuzzy-set-based scales to develop traditional scales that match natural responses. Rather, a fuzzy-set-based scale would have to be used to capture the varying degrees of imprecision in response across respondents. Other examples include the use of a range as a response to a particular item such as when estimating a price or a probability of purchase or the frequency of an event. The measurement of qualitative phenomena such as race when a sizable proportion of respondents have a mixed racial background may also be better accomplished by a fuzzy-set-based scale. Another possible application relates to questions which provide respondents with a list of factors, such as "Why do you go to the grocery store?" and asks them to circle all that apply. However, each alternative may apply to varying degrees and the response may be better captured by a fuzzy-set-based scale. The methodology may also offer interviewers a way of coding responses that are inherently imprecise.

Such information, which represents the *spread* or *range* of an individual's response to an item, cannot be collected completely with the

use of traditional scales, but may be critical as input for further analyses. Its incorporation into data analyses may increase the explained variance of results, because it represents a form of variation in individual response that is not captured by traditional analyses. It should be noted that, in these instances, the spread in response is not due to mismatch or to lack of certainty in response, both of which are aspects that can be diagnosed during scale development. Rather, this type of spread represents a characteristic of the content domain that should be captured by the measurement procedure.

Costs and Benefits of Using Fuzzy-Set-Based Scales. The use of the fuzzy-set-based methodology offers several benefits that are described above. However, it comes with additional costs when compared to traditional scales. Although data on items such as ease of usage provided some evidence that respondents were comfortable with using the fuzzy-set-based scale, it clearly involves more laborious data collection as well as data analysis procedures. Consequently, it is important to assess the costs and benefits of using this methodology in a particular situation. In this regard, the benefits of the proposed methodology may outweigh its costs when it is used only during scale development as a means of pretesting response scales and questionnaires, that is, the first two areas of applications described above, especially when the costs associated with imperfect measurement are large. These benefits may become more apparent with further research that refines the methodology and tests specific applications, as discussed subsequently.

In terms of the third area of applications described above pertaining to the measurement of imprecise phenomena, the use of this methodology may involve sizable costs. This is because the applications involve a substitution of traditional scales with fuzzy-set-based scales in all phases of data collection. Also, to a greater degree than in other applications, further research is necessary to develop and assess data analyses procedures that would have to be used for data from fuzzy-set-based scales. Even so, the methodology may be viable in situations where the use of the traditional scale is inappropriate, and it is important to capture varying degrees of imprecision in response. In broad terms, the proposed conceptualization and methodology offers researchers an approach for measuring phenomena that may heretofore have been considered too vague or ambiguous.

Implications for Future Research

Lines of future research should focus on ways of improving and implementing the fuzzy set methodology as well as on measurement issues

that can be studied with the use of the proposed approach. In terms of the methodology, variations on the instructions provided to respondents and membership levels used here need to be studied in order to make the task easier to complete and meaningful to respondents. Also, alternate ways of eliciting membership values should be examined. For example, one could use a graphic rating scale (cf. Griggs, 1980) that is divided into several labeled categories. Respondents are asked to indicate or mark off a range rather than a single point that may overlap with more than one category.

As described above, three alternate premises regarding spread in response have implications for alternate streams of research. If spread in response is viewed as mismatch, future research could examine different scales such as attribute rating scales, scales of attitude, scales of purchase intention, and nominal scales in terms of their number of response categories and category descriptors. The degree to which scales developed with the use of the fuzzy-set-based methodology reflect natural responses when compared to other scales in measuring overall liking and purchase intentions can also be studied. An issue of relevance that may fit into this line of research is the gauging of metric properties of responses. Understanding whether responses to scales are at nominal, ordinal, or interval levels would facilitate the design of appropriate scales. If spread in response is viewed as lack of certainty, research could examine the effects of question wording and content on responses to fuzzy-set-based scales. If spread in response is viewed as an inherent characteristic of certain content domains, future research could examine the performance of fuzzy-set-based scales versus traditional scales in measuring relevant phenomena, such as in predictive ability. In this regard, future research should also address alternate ways of analyzing data based on fuzzy-set-based scales. Although the present research was restricted to certain basic indicators of fuzziness, several indicators of intercategory relations from fuzzy set theory can be used to analyze data from fuzzy-set-based scales, such as measures of overlap between categories and inclusion of one category within another (cf. Smithson, 1982b).

In conclusion, information about the membership of responses in more than one response category cannot be inferred from existing measurement procedures that use single-category responses. Several important insights into scale response can be gained by conceptualizing response categories as fuzzy sets and broadening the existing perspective that scale responses can be captured by a single response category. Such a conceptualization allows for the application of techniques from fuzzy set theory to a broad range of measurement issues in marketing.

APPENDIX

Sample Scale and Instructions

Consider your response to the following question.

"How many hours of TV do you watch daily?"

Up to ½ hour	½ to 1 hour	1 to 1½ hours	1½ to 2 hours	More than 2 hours

If you think that your response is "Completely described by an alternative," you can circle the value "1.0" above that alternative in the scale shown below.

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
Up to ½ hour	½ to 1 hour	1 to 1½ hours	1½ to 2 hours	More than 2 hours

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").

REFERENCES

Alba, J. W., & Hutchinson, J. (1987). Dimensions of consumer expertise. *Journal of Consumer Research, 13*(4), 411–454.

Cox, E. (1980). The optimal number of response alternatives in a scale: A review. *Journal of Marketing Research, 17*, 407–422.

Griggs, A. O. (1980). Some problems concerning the use of rating scales for visual assessment. *Journal of the Market Research Society, 22*, 29–43.

Hunt, S., Sparkman, R., & Wilcox, J. (1982). The pretest in survey research: Issues and preliminary findings. *Journal of Marketing Research, 19*, 269–273.

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

Kaufman, A. (1975). *Theory of fuzzy sets* (Vol. 1). New York: Academic Press.
Kempton, W. (1978). Category grading and taxonomic relations: A mug is a sort of a cup. *American Ethnologist, 5*, 44–65.
Kempton, W. (1984). Interview methods for eliciting fuzzy categories. *Fuzzy Sets and Systems, 14*, 43–64.
Komorita, S. S., & Graham, W. K. (1965). Number of scale points and the reliability of scales. *Educational and Psychological Measurement, 25*(4), 987–995.
Lakoff, G. (1973). Hedges; a study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophic Logic, 2*, 458–508.
Lehmann, D. R., and Hulbert, J. (1972). Are Three-Point Scales Always Good Enough? *Journal of Marketing Research,* 9 (Nov.), 444–446.
McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets. *Memory and Cognition, 6*(4), 462–472.
Miller, G. A. (1956). The magical seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.
Park, C. W., & Lessig, V. P. (1981). Familiarity and its impact on consumer decision biases and heuristics. *Journal of Consumer Research, 8*, 223–230.
Perri, M., & Wolfgang, A. P. (1988). A modified measure of need for cognition. *Psychological Reports, 62*, 955–957.
Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. M. Moore (Ed.), *Cognitive development and the acquisition of language.* New York: Academic Press.
Smithson, M. (1982a). Models for fuzzy nominal data. *Theory and Decision, 14*, 51–74.
Smithson, M. (1982b). Applications of fuzzy set concepts to behavioral sciences. *Mathematical Social Sciences, 2*, 257–274.
Viswanathan, M., & Sudman, S. (1995). *Meaningful discrimination in scale responses: A perspective on the optimal number of response categories to use in a scale.* (Faculty Working Paper No. 92-0136), Urbana-Champaign: University of Illinois, Bureau of Economic and Business Research.
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.

The authors thank Seymour Sudman for his helpful comments, and especially thank Rajan Nataraajan for his detailed comments and ideas on an earlier draft of this article.

Correspondence regarding this article should be sent to: Madhubalan Viswanathan, Department of Business Administration, University of Illinois, 1206 South Sixth Street, Champaign, IL 61820 (e-mail: mviswana@ux1.cso.uiuc.edu).