A Paradigm for Validating Membership Functions

Paul T Dunn

Department of Computer Science
State University of California at San Francisco
1600 Holloway Avenue
San Francisco, California 94132

Abstract

Extensive experiments have been conducted by cognitive psychologists to verify models of human categorization and concept formation processes. These experiments frequently employ response time measurements to identify poor members and poor nonmembers of a given category. This paper examines the applicability of the same response time methodology to the problem of verifying fuzzy membership functions. Experimentally obtained response time measurements were found to correlate well with empirically derived fuzzy set membership functions. The general applicability of response time methodology to the problem of membership function verification is discussed.

Introduction

In their book "Fuzzy Sets and Systems: Theory and Applications", Dubois and Prade (1980) title the first chapter with a question - "Where Do "They" Come From ?"; "they" of course being membership functions. A multitude of approaches have been suggested and often implemented to answer the question. Exemplification (Zadeh 1972), various statistical methods (Hersh and Caramazza 1976, Civanlar and Trussell 1986), relative preferences (Saaty 1974) and parametric methods (Kuz'min 1981) to enumerate a few approaches. Each of these approaches has its own unique technical and philosophical characteristics, the illumination of which will be left to others.

Now that a number of fuzzy set researchers are strongly on the scent of generating membership functions, a new question comes to mind - How Do We Know "They" Are Realistic ? Fuzzy sets have been advanced as a mathematical formalism for vagueness, and vagueness is a topic which falls under the rubric of concept formation and categorization in cognitive science. Vagueness exists because concepts and categories do not have clear boundries. A realistic fuzzy set is therefore one which is consistent with the theoretical basis of its raison d'etre, human categorization and concept formation processes. A fuzzy set inconsistent with the underlying cognitive process calls to mind Marcellus's famous quote in

Hamlet that "something is rotten in the state of Denmark." We turn now to consider cognitive science's understanding of categorization and concept formation.

Models of categorization and concept formation fall broadly into two classes, semantic network models and set theoretic models. Collins and Quillian's (1969) model depicted in figure 1.1 provides a overview of the principals underlying most semantic network models. Such models assume that concepts exist as nodes on a network connected by relations such as an arabian "is a" horse. Verification of the proposition "an arabian is a horse" requires a comparison of the network's relation between arabian and horse with that asserted in the proposition. Semantic network models address issues somewhat different from those we are interested in and we will not consider them further.

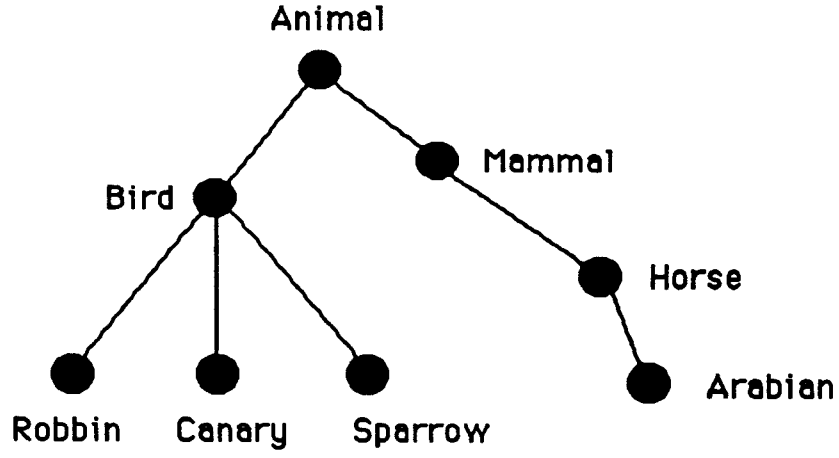

Figure 1.1. Collin's and Quillian's Semantic Network Model.

Three theories account for most of the set theoretic models: (a) prototype theory, (b) feature comparison and (c) instance comparison. Before considering these theories in greater detail we will review the typicality effects from which they draw most of their empirical support. More typical category instances: (1) are learned before less typical instances, (2) take less time to verify as category members than less typical instances, (3) are named more frequently when subjects are asked for examples of a given category, (4) are identified as anchor points in linguistic hedges and (5) are reliably rated by subjects as being more typical.

Rosch (1973) argued that most natural categories have an internal structure which is not comprised of equivalent undifferentiated instances. Categories have a "core meaning" composed of the category's "clearest cases" (prototypes) which is surrounded by increasingly less typical instances of the "core meaning." In Rosch's view natural categories do not have clear boundries.

Rosch (1973) hypothesized that the time for subjects to respond affirmatively to statements like "an X is a Y" (where X is an instance and Y

is a category) would be less for central category members than for peripheral category members. To explore this hypothesis Rosch presented subjects with instances (e.g. pear) of various categories (e.g. fruit). Subjects rated the instance as a very good fit to a very poor fit by checking one of seven blank spaces. From this pool of ratings Rosch generated four kinds of instance-category pairs, false central, false peripheral, true central and true peripheral. The time for subjects to respond true or false to examples of each kind of instance-category pair was measured. Subjects took less time to respond to central instance category pairs than to their corresponding peripheral instance category pairs.

Prototype theory describes the content and structure of categories and not the actual representation process itself which may be based on images, feature lists or structural descriptions. All that is required of processing models is that they not violate the known behavior of prototypes. Thus a processing model should not produce shorter category verification times for poor members than for good members.

Smith, Shoben and Rips (1974) believed that categorization was based on a feature comparison process. They argued that a term's meaning could be represented by a combination of defining features and characteristic features. Defining features are those on the term's feature continuum which are more essential to the term's meaning. Characteristic features are more accidental and contribute less to the term's meaning. In general the response time depends on the degree of similarity between the instance and the category.

Brooks (1978) demonstrated that subjects are often not able to explicitly state defining attributes and categorization rules. His experimental subjects were able to develop concepts unconsciously and nonanalytically. When learning pairs of letter strings with cities or animals (e.g. MRRRRRM - bison, VVTRXRR - Paris) they were able to correctly classify instances as old world or new world. Brooks believed that instances were identified as category members by their global similarity to known category members. A flower would be categorized as such because it vaguely resembled something like that seen previously. Medin and Smith (1981) advanced an instance comparison model similar to Brooks'.

Figure 1.2 provides an overview of how one would categorize a house as belonging to the set of nice houses based on each of reviewed categorization theories.

Model	Categorization Question(s)
Feature Comparison	How do the features of this house relate to the features of the category of nice houses ? For example, does it have a swimming pool ?
Instance Comparison	How does this house compare with other nice houses I have seen ?
Prototype Theory	How does this house compare with the nice house prototype(s) ?

Figure 1.2. Categorizing a house as a nice house.

In practice, RTs are not as deterministic as might be suggested by the aforementioned categorization theories. Subjects may employ a variety of heuristics for responding true or false to a series of phases of the form "an 'instance' is a 'category'"; heuristics which cannot be employed in single instance real life situations. What is important is the general trend of RT in categorization experiments: faster responses for a category's clear members and nonmembers than for its peripheral members and nonmembers.

With a theoretical basis for understanding typicality effects and how they can identify central and peripheral category members, we are free to explore their use in verifying membership functions. In particular we will concentrate on RT measurements and expect higher RTs where $0 < \mu < 1$ than where $\mu = 0$ or 1.

Experiment

Response time behavior to agree or disagree with statements of the form X (an instance) is Y (a category) has been studied extensively in categorization and concept formation experiments. The time to agree or disagree is consistently higher for poor members and poor nonmembers of a category than for clear members and clear nonmembers. Accordingly response time peaks should occur in the vague region of fuzzy sets where $0 < \mu_Y(x) < 1$ and can been seen in a general sense as a validation method for membership functions. Response time data was gathered to demonstrate the principle of membership function verification with membership functions derived from subjects' agree/disagree responses.

Method:

Subjects. Fourteen unpaid mathematics and computer science volunteers at the State University of California at San Francisco served as subjects. Half of the subjects were at least generally familiar with the theory of fuzzy sets and half reported no prior exposure to fuzzy set theory.

Stimuli. Phrases of the form X is Y on a computer screen were used as the stimuli. X ranged in value from 54 degrees fahrenheit to 86 degrees fahrenheit outside air temperature graduated in one degree increments. The linguistic variable Y took on the values cool, warm and hot for each value of X.

Procedure. Subjects were run individually. The test phrases were shown on the screen of a Macintosh computer and the response time for subjects to agree or disagree with each phrase was recorded. Below each phrase on the computer screen was an agree button and a disagree button for responding to the phrase. Buttons were selected using the Macintosh's mouse. After responding to a phrase, subjects pushed the next button to view a new phrase. Use of the next button forced subjects to center the mouse between agree/disagree responses and provided subjects with an opportunity to rest between phrases as necessary. A sample test phrase is shown in Figure 2.1.

Subjects responded to a randomized set of phrases and to a nonrandomized set of phrases. For the nonrandomized phrases X was monotonically increasing from 54 degrees to 86 degrees for each of the linguistic Y variables cool, warm and hot. Half of the subjects responded to the randomized phrases first and half of the subjects responded to the nonrandomized phrases first. The range of X and Y values was provided to subjects before beginning the experiment. All subjects participated in a short practice run with height data prior to commencing their timed temperature runs. RT was collected in seconds and is reported throughout this paper in seconds.

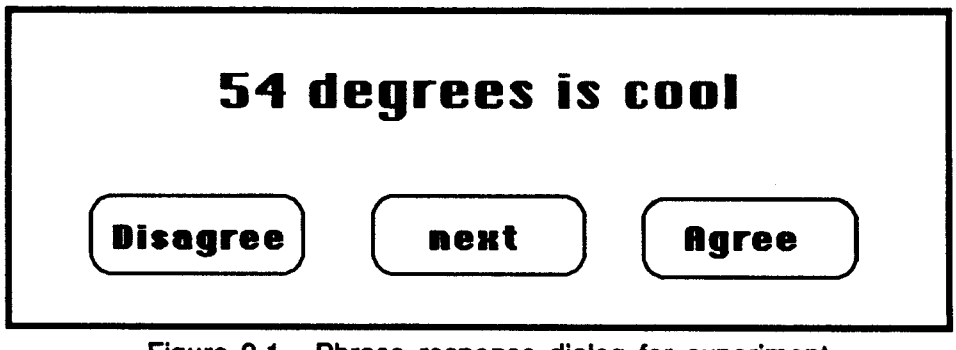

Figure 2.1. Phrase response dialog for experiment.

Discussion of Results

Raw RT data from a number of the subjects required adjustment to address two problems. Several of the subjects reported the randomized run to be a "mind numbing" experience and despite admonitions to the contrary engaged in talking and space gazing. RT measurements which were known to be corrupted in this way were replaced with the average value of the two adjacent RT measurements after the data was monotonically ordered. A second source of difficulty arose from obviously incorrect agree/disagree responses in the randomized runs. These discrepancies were addressed by not recognizing changes in the agree/disagree variable unless two successive same values were obtained.

During the nonrandomized runs subjects appeared to slowly decrease their RT for a given linguistic variable Y. When Y changed response time increased and subsequently began decreasing again. Decreasing response time behavior for a given Y is generally consistent with learning of a repetitive task. From visual observation it was not obvious whether there was a RT learning effect for the randomized runs.

Figure 3.1 depicts S1's nonrandomized RT data for all values of X for Y = cool. S1's randomized RT data in figure 3.2 for the 99 XY pairs from all three Y values depicts the generally higher standard deviation for randomized RT data noted in figure 3.3. Data in figure 3.2 is shown in randomized order as presented to S1.

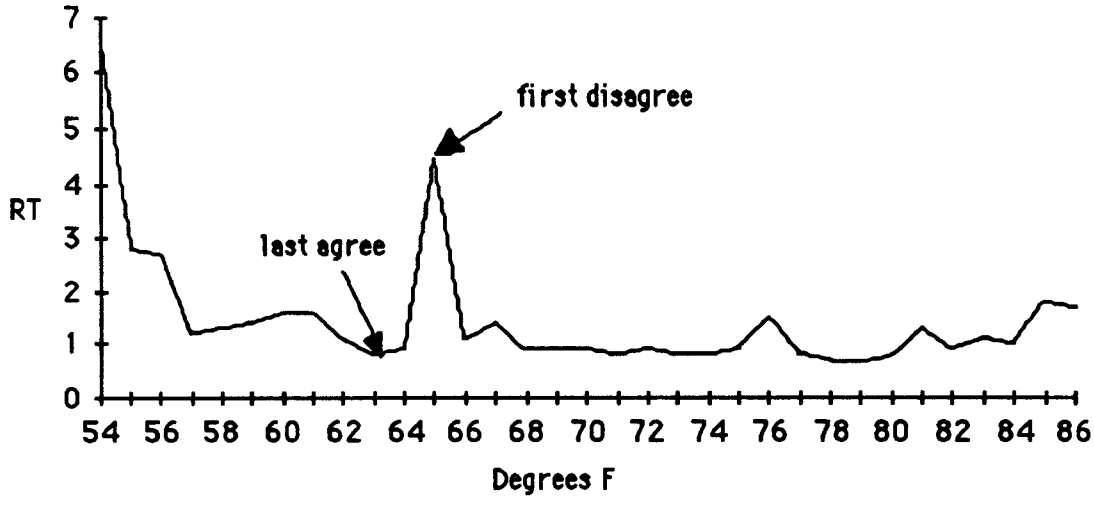

Figure 3.1. Unsmoothed Nonrandom RT Data for Y = Cool: Subject 1

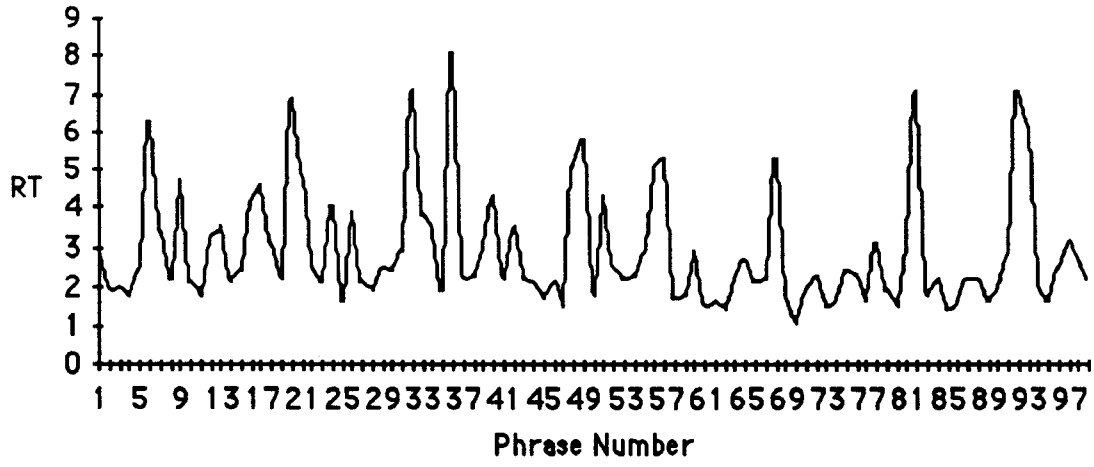

Figure 3.2. Unsmoothed random RT Data: Subject 1

A linear regression model was constructed for the experiment's RT data. Predicted RT served as the dependent variable. Of the five independent variables four were indicator variables related to the agree/disagree responses and the fifth was a decreasing function intended to model the observed RT learning effect. The inflection point is defined as the imaginary value of X between a change in a subject's agree/disagree response for a given Y.

$$PRT = a + bI_1 + b_2I_2 + b_3I_3 + b_4I_4 + cL \qquad (1)$$

PRT: predicted response time
a: y intercept
b_1-b_4: indicator variable coefficients
p: imaginary inflection point phrase
I_1: indicator variable at inflection point p-2
I_2: indicator variable at inflection point p-1
I_3: indicator variable at inflection point p+1
I_4: indicator variable at inflection point p+2
c: learning effect variable's coefficient
L: learning effect compensation variable, defined as ln(k+m) where K > 0 and m is the phrase number.

R-squared values were highest for the nonrandomized runs when a model was constructed for each group of 33 responses for a given Y rather than the 99 responses encompassing all three Y values. In addition to a need for mouse coordination learning for a given run there appeared to be a need for context learning as the subject considered what a cool day felt like. Lack of context learning in the randomized environment may explain the need for reinitialization of the learning variable for each new Y for the nonrandomized runs and why the learning variable coefficient c was significant with a 97.5% confidence limit for only two subjects. Figures

3.3 and 3.4 list the variables which were significant with a 97.5% confidence limit for the nonrandomized runs and for the randomized runs.

Subject	Y = Cool Variable	R-squared	Y = Warm Variable	R-squared	Y = Hot Variable	R-squared
S1	I3,L	0.87	I2,I3,L	0.54	I3,L	0.88
S2	I2,I3,L	0.68	I2,I3,L	0.80	I3	0.48
S3	I3,L	0.83	I3,L	0.58	L	0.35
S4	I1,I2,I4,L	0.70	I3,L	0.43	I3,L	0.63
S5	I3,L	0.58	I1,I3,L	0.48	I2,L	0.40
S6	I3,L	0.70	I3,L	0.67	I3,L	0.40
S7	I2,I3,L	0.77	I2,I3,L	0.47	I1-I3,L	0.58
S8	L	0.31	I1-I3,L	0.56	I1-I3,L	0.77
S9	I1,I3,L	0.64	L	0.31	NA	NA
S10	I3,L	0.53	none	0.05	I3,L	0.55
S11	NA	NA	I2,L	0.35	none	0.05
S12	L	0.44	I2	0.16	I2,L	0.55
S13	L	0.58	I1,I2	0.50	I2,L	0.28
S14	I3,L	0.85	I3,L	0.89	I2,I3	0.61

NOTE: Coefficients significant with a 97.5% confidence limit.

Figure 3.3. Significant Coefficients and R-squared values for nonrandomized runs.

Subject	Variable	R-squared
S1	I2,I3	0.18
S2	I1,I2	0.11
S3	I1,L	0.20
S4	I3,I4	0.08
S5	I2,I3	0.11
S6	I4	0.10
S7	none	NA
S8	none	NA
S9	I1,I2	0.12
S10	I1,I3,I4	0.26
S11	I4	0.05
S12	I2	0.05
S13	I1,I2	0.11
S14	I1,I2,L	0.26

NOTE: Coefficients significant with a 97.5% confidence limit.

Figure 3.4. Significant Coefficients and R-squared values for randomized runs.

The results of the experiment are consistent with the prior findings of categorization and concept formation experiments. RTs were greater for marginal members and marginal nonmembers of a category than for clear members and clear nonmembers. Although the RT data is too noisy for constructing membership functions, the data appears well suited for

identifying the general location of a fuzzy set's vague region.

The greater RT for poor members and poor nonmembers of a category in concept formation and categorization experiments should translate into RT peaks in a membership function's vague region, or where $0 < \mu_Y(x) < 1$. In this way RT methodology can be used to validate the general shape of a given membership function. We turn now to examine this claim by comparing the experiment's RT data with several experimentally determined membership functions obtained from subject agree/disagree responses.

Zadeh (1968) extended the relation between an event Y's probability (P) and the expected value (E) of its membership function μ_Y. Specifically:

$$P_Y(X) = E(\mu_Y(X)). \qquad (2)$$

Hersh and Caramazza (1976) used equation (2) to construct membership functions for size descriptors of black squares projected on a white background. The binary responses of Hersh and Caramazza's subjects were averaged to generate membership functions for the available square size descriptors.

Normalized membership functions for the temperature descriptors were generated using equation (2) and the agree/disagree data from the RT experiment. Thus:

$$\mu_Y(x) = \text{Normalized } ((\Sigma\ AD_i)/n) \text{ for } i=1 \text{ to } i=n \qquad (3)$$

where AD_i is the agree/disagree response for subject i and $AD_i = 1$ if the subject agrees and $AD_i = 0$ if the subject disagrees and n is the number of subjects.

In order to compare RT peaks with membership function values the RT data was normalized by subject and averaged. Normalization of the RT data by subject prevented data from subjects with high mean RTs from obscuring the RT data relationships from subjects with low mean RTs. The aggregated RT (ART) for a given X in Y was calculated as:

$$ART_Y(x) = \Sigma \text{ normalized } RT_i \text{ for all } i \qquad (4)$$

The data from subjects exhibiting entailment of warm and hot agree/disagree responses were eliminated from the ART and μ calculations for Y = warm for the nonrandom data and for Y = cool, warm and hot for the random data. As with the individual subject data, the ART data contained a learning trend component. This learning trend component was estimated by regressing ART with one of three decreasing functions listed in Figure 3.5. To depict the relationship between ART and μ the predicted ART (PART) was subtracted from the ART to provide the ART residuals (R) for a given Y for the nonrandom data or for all Ys for the random data:

$$PART(x) = a + b^*L_j(x) \qquad (5)$$

$$R(x) = ART(x) - PART(x) \qquad (6)$$

i	Learning Function L
1	1/ln (k+m)
2	exp(-km)
3	((N-1)/(1-N))m+(N+1)

k : experimentally determined
m : phrase number in data group
N : number of phrases in this data group

Figure 3.5. PART Learning Functions.

As expected positive values of the residual occurred most commonly in a fuzzy set's vague region. An intuitive feeling for the distribution of positive values of R as a function of μ_Y can be obtained from figures 3.6 and 3.7. The thick black horizontal line marks the point below which $R < 0$ and above which $R > 0$.

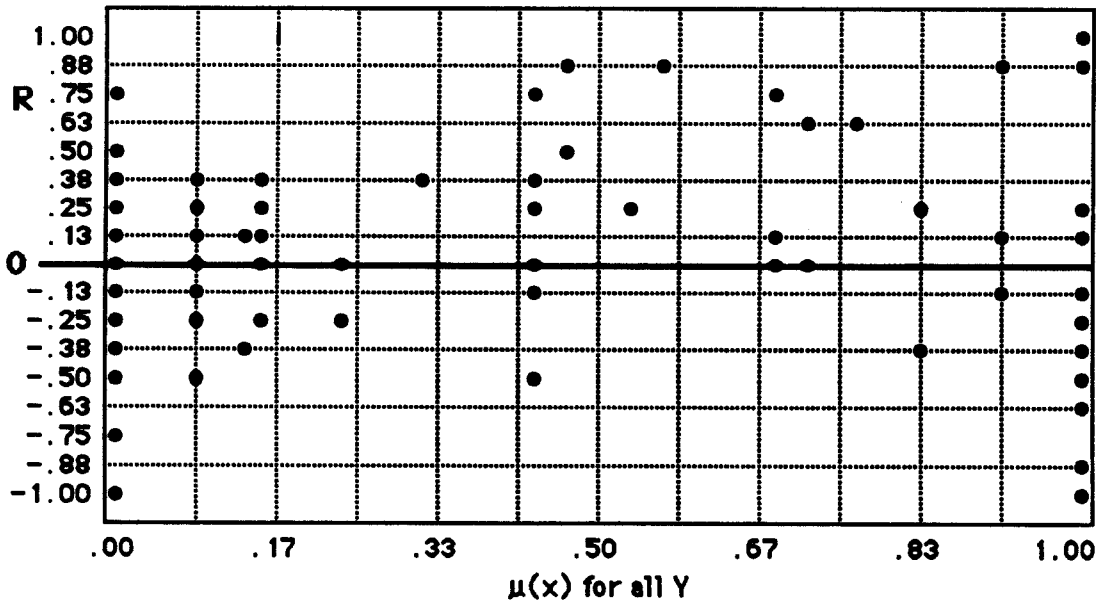

Figure 3.6. The residual R vs μ_Y for all Ys for nonrandomized data.

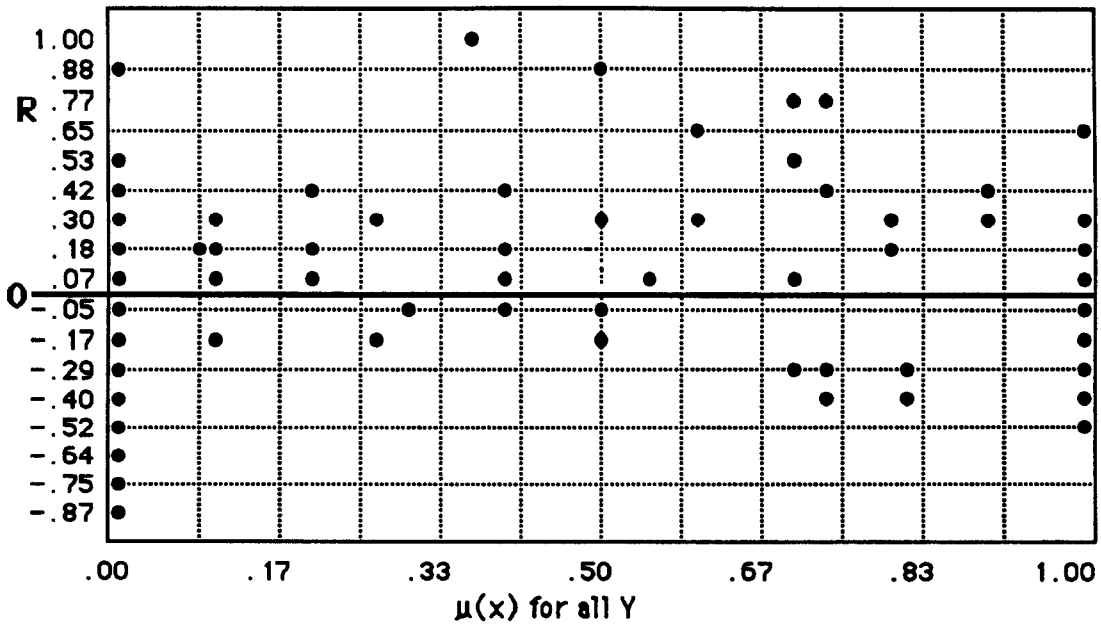

Figure 3.7. The residual R vs μ_Y for all Ys for randomized data.

Summary

There are differences in the membership functions obtained from nonrandomized and randomized phrases. Although these differences relate to the value of μ_Y for a given X and not to the general behavior of RT data, they do raise a question. Should one use randomized or nonrandomized supports when considering the performance of a membership function generation technique ? Without additional experimentation this is not a straight forward question to answer. Although biases can result from the use of nonrandomized data, individuals in real life situations undoubtedly form categories from nonrandomized data. For example, real world ordering of purchase alternatives on price can produce categorization based on nonrandomized data.

Although RT methodology appears overly time consuming for daily verification of membership functions, it has great value in examining the general validity of a particular membership function generation technique. Membership function generation techniques significantly at odds with the cognitive psychology underlying the categorization process raise serious questions about their validity. RT measurement's theoretical foundation in cognitive psychology can be seen as providing a valuable tool for verifying the overall shape of membership functions.

References

Brooks, L. Nonanalytic Concept Formation and Memory for Instances. In E. Rosch and B.B. Lloyd (eds), *Cognition and Categorization.* Hillsdale New Jersey, Lawrence Erlbaum Associates, 1978.

Civanlar, H.R. and Trussell, H.J. "Determination and Applications of Membership Functions Based On Statistical Data." In *Recent Developments in the Theory and Applications of Fuzzy Sets, Proceedings of NAFIPS' 86.* New Orleans, 1986.

Collins, A.M. and Quillian, M.R. "Retrieval Time from Semantic Memory." *Journal of Verbal Learning and Verbal Behaviour*, 1969, **8**, pp. 240-247.

Dubois, D. and Prade, H. (1980) Fuzzy Sets and Systems: Theory and Applications. New York, Academic Press.

Hersh, H.M. and Caramazza, A. "A Fuzzy Set Approach to Modifiers and Vagueness in Natural Language." *Journal of Experimental Psychology,* 1976, **105**, pp. 254-276.

Kuz'min, V.B. "A Parametric Approach to Description of Linguistic Values of Variables and Hedges." *Fuzzy Sets and Systems*, 1981, **6**, pp. 27-41.

Medin, D.L. and Smith, E.E. "Strategies and Classification Learning." *Journal of Experimental Psychology: Human Learning and Memory*, 1981, **4**, pp. 241-253.

Rosch, E. On the Internal Structure of Perceptual and Semantic Categories. In T.E. Moore (ed), *Cognitive Development and the Acquisition of Language.* 1973.

Saaty, T.L. "Measuring the Fuzziness of Sets." *Journal of Cybernetics*, 1974, **4**, pp. 53-61.

Smith, E.E., Shoben, E.J., and Rips, L.J. "Structure and Process in Semantic Memory: A Featural Model for Semantic Memory." *Psychological Review*, 1974, **81**, pp. 214-241.

Zadeh, L.A. "Probability Measures of Fuzzy Events." *Journal of Mathematical Analysis and Applications*, 1968, **23**, pp. 421-427.

Zadeh, L.A. "Fuzzy Languages and their Relations to Human and Machine Intelligence." From *Man Computer Processing International Conference, Bordeaux 1970.* 1972, S. Karger, Basel pp. 130-165.