Cue Preference in a Multidimensional Categorization Task

Patricia M. Berretty (berretty@fordham.edu)
Department of Psychology,
Bronx, NY 10458-5198 USA

Abstract

Many natural categories vary along multiple dimensions. The present studies address two main questions underlying categorization with multiple dimensions. First, how well can humans perform in a categorization task consisting of five categories varying along nine continuously valued dimensions? Second, what are the properties of the cues preferred by humans if not all the available cues are used? Remarkably, participants not only learned to distinguish among the five categories, but they also learned to do so using only the relevant dimensions. A satisficing model of categorization was best able to account for participants' responses. In addition, in a cue preference task, the results showed that nearly all participants preferred to use the dimension with the greatest variance when the number of dimensions available was restricted, in accord with predictions made by the satisficing model.

Introduction

Categorization has been studied by many disciplines including psychology and machine learning. In the area of psychology, the psychological processes underlying human categorization have been investigated. One common approach to determining these processes has been to teach humans to learn novel categories based on very simple stimuli that vary along only a few dimensions. In such simple situations, the complex calculations involved in some of the popular models of categorization (e.g., Nosofsky's (1986) generalized context model; Ashby's (Ashby & Gott, 1988; Ashby & Perrin, 1988) decision bound theory) may be psychologically plausible. However, the results of these experiments are then assumed to be generalizable to categories whose members vary along many dimensions. It seems unreasonable to assume that humans are capable of the even more complex calculations required with an increase in category dimensionality. For example, Nosofsky, Palmeri, and McKinley (1994) "question the plausibility of exemplar storage processes and the vast memory resources that they seem to require" (p.53).

Machine learning, on the other hand, has been primarily concerned with developing algorithms based on experts in specific domains (Quinlan, 1986) – although the algorithms themselves tend to be general-purpose algorithms (i.e., the algorithms are intended to apply to any categorization task). These algorithms have been developed using large data sets that vary along many dimensions. Therefore, an important step in such algorithms is determining which dimensions from the set of possible dimensions should be used. However, the different methods used to model this step usually involve complex computations and thus are also not psychologically plausible.

What follows is a brief review of two popular categorization models (exemplar models and decision bound theory), as well as a review of a satisficing model of categorization (categorization by elimination). Next, a multi-dimension, multi-category task is described, including a discussion of how well the above three models can account for human responses in such a task. The paper concludes with a brief discussion on the learning of relevant cues in the multi-dimension, multi-category task.

Review of Models

Exemplar Models

Exemplar models (Brooks, 1978; Estes, 1986; Medin & Schaffer, 1978; Nosofsky, 1986) assume that when presented with a novel object, humans compute the similarity between that object and all exemplars of every category in which the novel object could be placed. In theory, the object is placed into the category with which it is most similar, however most exemplar models assume probability matching. Nosofsky's (1986) generalized context model (GCM) allows for variation in the amount of attention given to different features during categorization (see also Medin & Schaffer, 1978). Therefore, it is possible that different cues will be used in different tasks. However, this attention weight remains the same for the entire stimulus set for each particular categorization task, rather than varying across different objects belonging to the same category. GCM uses a probabilistic response rule based on the Luce-Shepard choice model. The probability of placing stimulus i into category j is computed by summing the similarity between stimulus i and all objects in category j along every dimension and then weighting the summed similarity by the bias to respond with category j. The weighted summed similarity is divided by the sum of the weighted summed similarity of stimulus i to each category. Similarity is usually either an exponential (for separable stimuli) or gaussian (for integral stimuli) function of psychological distance (Shepard, 1964). Psychological distance is computed by the Minkowski r-metric with the addition of two parameters, c and w_k, where c is the discriminability parameter which takes into account that stimuli will look more distinct as experience is gained and w_k is the attention weight for the kth dimension.

Decision Bound Theory

Decision Bound Theory (or DBT—see Ashby & Gott, 1988) assumes that there is a multidimensional region associated with each category, and therefore, that categories are separated by bounds. DBT uses a deterministic response rule. An object is categorized according to the region of perceptual space in which it lies. The perceptual space is divided into regions by decision bounds. For two categories (A and B) each composed of two dimensions (x and y), an object will be placed into category A if the estimated likelihood ratio is greater than some bias, where the likelihood ratio refers to the ratio between the likelihood that stimulus i comes from category A and the likelihood that stimulus i comes from category B. The parameters of this model are b, a response bias; a mean and variance for each dimension (which are usually absorbed into the bound parameters); correlations between pairs of dimensions; as well as parameters to define the decision bound.

Both of these psychological models categorize by integrating cues and using all the cues available (except in exemplar models if a cue has an attention weight of zero). But the memory requirements of these models do differ. GCM assumes that all exemplars ever encountered are stored and used when categorizing a novel object, while DBT only needs to store the bound-determining parameters of each category. In comparison, the Categorization by Elimination algorithm (described below) typically requires as little memory as DBT but it does not integrate all available cues.

Categorization by Elimination

Categorization by Elimination (CBE) was originally developed to describe people's categorization judgments in an animate motion task (see Blythe, Miller, & Todd, 1996). CBE is closely related to Tversky's (1972) Elimination by Aspects model of choice. CBE is a noncompensatory model of categorization, in that it uses cues in a particular order, and categorization decisions made by earlier cues cannot be altered (or compensated for) by later cues. In CBE, cues are ordered and used according to their probability of success. For the present purpose probability of success is defined as a measure of how accurately a single cue categorizes some set of stimuli (i.e., percent correct). This is calculated by running CBE only using the single cue in question, and seeing how many correct categorizations the algorithm is able to make. (If using the single cue results in CBE being unable to decide between multiple categories for a particular stimulus, as will often be the case, the algorithm chooses one of those categories at random—in this case, probability of success will be related to a cue's discriminatory power.)

CBE assumes that cue values are divided up into bins (either nominal or continuous) which correspond to certain categories. To build up the appropriate bin structures, the relevant cue dimensions to use must be determined ahead of time. At present, complete bin structures are constructed before testing CBE's categorization performance. Bins can be constructed in a variety of ways from the training examples by determining low and high cue value boundaries for each category on each dimension. These boundaries are then used to divide up each dimension into the cue-value ranges that form the bins. Thus, CBE only needs to store two values per category per cue dimension, independent of the number of objects encountered.

Categorization with Multiple Dimensions

The majority of psychological studies of categorization have used simple artificial stimuli (e.g., semicircles in two-dimensional space—Nosofsky, 1986) that vary on only a few (two to four) dimensions[1]. This is in contrast to the more natural high-dimensionality machine learning applications, such as wine tasting (Aeberhard, Coomans, & Devel, 1994) or handwriting recognition (Martin & Pittman, 1991). It remains to be demonstrated how optimal humans can be when categorizing objects using many continuously valued dimensions. In addition, the predominant psychological models of categorization have not addressed the issue of how people can categorize a multidimensional object when they are constrained by limited information.

Berretty and her colleagues (Berretty, Todd, & Martignon, 1999; Berretty, Todd, & Blythe, 1997) have illustrated that it is possible for a satisficing model that does not use all the available cues to categorize objects, to perform comparably to integrative models on natural data sets. The purpose of the first experiment in this paper is to investigate whether such a satisficing model is able to account for human categorization data from a multi-dimensional, multi-category task. In Experiment 1a, humans are trained to learn categories that vary along nine dimensions. The generalized context model, categorization by elimination, and a form of decision bound theory will be tested to determine how well each model fits the participants' responses. The purpose of the second experiment is to determine how well humans

[1] Posner and Keele (1968) have used random dot stimuli to test human classification, however, the number of dimensions is indeterminable.

are able to categorize when information is limited. In addition, Experiment 1b investigates the properties of the dimensions people prefer to use when information is limited.

Participants

Four graduate students from the University of California, Santa Barbara participated in Experiment 1a and 1b. All participants had normal or corrected vision. Each participant was paid $8 per hour.

Method

Design The design consisted of five different categories that vary along nine dimensions, where only three of the dimensions are necessary for accurate categorization. The values for each category were generated from a multivariate normal distribution where variance(dim 1) > variance(dim 3) > variance(dim 2), with the variance for the remaining 6 dimensions equal to the variance along dimension 3. All uni-dimensional rules that best separate two categories with the same mean on the other two relevant dimensions have an accuracy of 90% (i.e., category overlap along each pair of dimensions was set to 10%).

Procedure Participants were told that they were to learn five different categories that were equally represented during each learning session. Participants were instructed that they may or may not need to use all the dimensions available to them. Participants were run over consecutive days until learning curves leveled off. Each day consisted of 20 blocks with 50 trials per block (for a total of 1000 trials per day). Stimulus display was response terminated and corrective feedback was given after every trial. Thus, if a subject responded 'A' to an exemplar from category B, a low tone sounded followed by a 'B' appearing on the screen. In addition, overall percent correct was given after every learning block.

A cue preference task (Experiment 1b) was administered to participants the day after learning ended. The cue preference day began with a practice block in which participants simply categorized 50 stimuli as they had done on previous days. The practice block was followed by twelve blocks, each consisting of 50 trials. Each trial began with the presentation of one of the three relevant dimensions. Participants then made a categorization judgment based on only that one dimension. After making a judgment, participants chose another dimension and then made another categorization judgment. Thus, two judgments were made for the same stimulus. The first judgment was based on only one experimenter-chosen dimension, while the second judgment was based on two dimensions. No feedback was given during the last twelve blocks of the test day.

Stimuli and Materials Stimuli were generated using the GRT Toolbox (Alfonso-Reese, 1995). Values along every dimension were transformed from number of dots per square into actual screen coordinates. Each dimension was represented as a texture in one of nine possible squares on a computer screen. The location of the three relevant dimensions was different for each subject with the constraint that the center square (in a 3x3 grid) will never be one of the relevant dimensions.
Stimuli were presented on a SuperMac Technology 17T Color Display driven by a PowerMacintosh G3 running a Psychophysics Toolbox (Brainard, 1997) and low-level VideoToolbox (Pelli, 1997) within MATLAB (The MathWorks, Inc., 1998). Each participant sat 18 inches from the monitor. The height of the center square of the stimuli was constrained such that visual angle was less than 2°.

Results and Discussion

Experiment 1A Learning for three of the four participants reached asymptote after five days, while the fourth participant required six days. Participants 1, 2, and 3 achieved an overall accuracy of approximately 70% by the last day, while Participant 4 only achieved an overall accuracy of approximately 60% on the last day. The optimal percent correct was 81.9%. Participants' responses for the last day (without the first block) were randomly split into two halves (training and testing sets) five times. Each split was constrained to contain approximately the same number of stimuli from each category.

The Categorization by Elimination algorithm, the Deterministic Generalized Context Model (see Ashby & Maddox, 1993), and six versions of Decision Bound Theory were fit to each participant's training set responses. For CBE, low and high values of each bin along each dimension, as well as the cue order, were estimated from the responses in the training set. The parameters estimated for GCM were the sensitivity parameter, an attention weight for each dimension, the bias towards each category, and the gamma parameter (which is a measure of response selection). For fitting the GCM, a Euclidean-Gaussian distance-similarity metric was used (see Maddox & Ashby, 1998).

The six versions of DBT were all Independent Decisions Classifiers, which is a special case of Decision Bound Theory in which each dimension is assumed to be independent of the other dimensions (see Ashby & Gott, 1988; Ashby & Maddox, 1990). This version of DBT was used since the best fitting bound (to separate the categories) is perpendicular to each of the three relevant dimensions. In the versions of the Independent Decisions Classifier tested here, one criterion is placed along one dimension. Two criteria are then placed along a second dimension and four criteria are placed along the third dimension. All

Table 1: AIC Scores for Experiment 1A

	P1 Train	P1 Test	P2 Train	P2 Test	P3 Train	P3 Test	P4 Train	P4 Test
GCM	585.4	633.6	739.42	823.08	647.33	687.14	814.4	835.24
DBT	594.74	638.16	742.63	780.87	645.32	665.22	809.55	824.54
CBE	646.28	643.59	638.32	640.36	624.5	634.86	656.04	646.85

possible combinations of the three relevant dimensions were tested.

As mentioned earlier, all three models were fit to part of the data set (the training set) and the best fitting parameters estimates were obtained. These parameters were then used to determine the models' accuracy on the remaining data (the testing set). A potential problem with multi-parameter models is that these models may be prone to overfit the data. That is, they actually fit the noise present in the data in order to achieve high accuracy. Training the model on a subset of the data and testing the model on the rest of the data may assess a model's "true" performance.

The AIC goodness-of-fit statistic was used to compare the fits of the three models.

$$AIC(M_i) = -2\ln L_i + 2v_i$$

Where $\ln L_I$ refers to the negative log likelihood value for model M_i obtained through maximum likelihood estimation and v_i refers to the number of free parameters in model M_i. The smaller the AIC score, the closer the model is to the "true" model (Ashby, 1992).

Goodness-of-fit values for each participant (averaged over the five training and five testing sets) are shown in Table 1. Each row corresponds to one of the three models while each column refers to each participant's training and testing sets. The generalized context model was best able to account for Participant 1's training and testing data. Categorization by elimination was best able to account for Participant 2, Participant 3, and Participant 4's training and testing data.

Experiment 1B Experiment 1b was designed to answer two questions. First, how well can humans perform in a categorization task when dimensionality is reduced? Second, what are the properties of the dimensions preferred by humans? Obviously, one of the most important features of a cue is how accurate that cue is in categorizing objects when used alone. Another property of cues is the range of values possible, that is, the variance of a cue. It seems reasonable to assume that humans are able to learn the accuracy of various cues and would use those cues that are more accurate. Given this assumption, all three of the relevant dimensions are equally accurate when used alone. However, the question of whether humans prefer to use cues with more or less variance is addressed by having different variances for the three relevant dimensions.

In Experiment 1b (conducted after performance asymptotes) participants were given one dimension and asked for a categorization judgment.[2] Participants then chose a second dimension (from the remaining eight dimensions) and made a categorization judgment based on only those two dimensions. Only the three relevant dimensions for the categorization task were used in Experiment 1b as the first cue presented to the participant. Both high and low values of these dimensions were given to the participants. Dimension values were selected from the categories such that the values were always less than (or greater than) the best fitting criteria values for that dimension (i.e., only dimensional values from nonoverlapping category regions were presented).

The first major result to notice from this experiment is the overall percent correct participants achieved, which is shown in Table 2. The optimal percent correct possible with only two categories is 51.6%. Participant 3 was very close to optimal, while Participants 2 and 4 actually performed better than would be expected. In addition, Participant 4 actually performed better in Experiment 1b than in Experiment 1a!

Table 2: Overall Percent Correct in Experiment 1B

	Participant 1	2	3	4
Percent Correct	42.67	55.23	49.83	64.5

The results from Experiment 1b indicate that participants did indeed learn which of the cues in Experiment 1a were relevant. All four participants chose (nearly always, if not always) one of the three relevant dimensions as their second cue in Experiment 1b (see Table 3). This indicates that participants were not using any of the other dimensions during Experiment 1a[3].

[2] Participants were given the first cue to insure that all three of the relevant dimensions would be chosen. If participants were allowed to choose the first cue to use, it is possible that the same cue would be used first for each trial.

[3] This does not rule out the possibility that participants were using other dimensions in Experiment 1a, but preferred to use one of the three relevant dimensions when limited in the number of dimensions available to them. However, verbal

Table 3: Dimension Preference for Participants 1-4

Dimension Presented	Dimension Chosen by Participant 1			
	1	2	3	4-9
1	23	150	25	0
2	188	9	2	0
3	186	11	0	0
	Dimension Chosen by Participant 2			
	1	2	3	4-9
1	9	80	103	1
2	86	3	100	5
3	91	88	7	7
	Dimension Chosen by Participant 3			
	1	2	3	4-9
1	16	162	22	0
2	162	5	27	0
3	186	9	4	0
	Dimension Chosen by Participant 4			
	1	2	3	4-9
1	15	45	134	0
2	113	0	87	0
3	133	59	8	0

According to CBE when dimension 1 is presented, dimension 3 should be chosen and when dimension 2 or 3 is presented, dimension 1 should be chosen. When dimension 1 was presented first two of the participants preferred the dimension with the highest probability of success (dimension 3). When dimension 2 was presented first, three of the participants preferred the dimension with the highest probability of success (dimension1). All four participants preferred the dimension with the highest probability of success (dimension 1) when dimension 3 was presented first.

Overall, the participants generally chose the second dimension in accord with predictions made by CBE.

Learning Relevant Cues

Given the difficulty of the task in Experiment 1a, it is remarkable that the participants were able to learn the relevant cues. As shown above, all four participants chose (nearly always, if not always) the three relevant dimensions as their second cue in Experiment 1b. But how did cue use progress as the participants learned the different categories in Experiment 1a? To answer this question three different versions of MDS were fit to the participants' category confusion matrices from each half of each day in order to determine how many cues were used by each participant for a particular data set. MDS_1 uses only one dimension, MDS_2 uses two dimensions, and MDS_3 uses three dimensions to account for the participants' confusions. A χ^2 analysis was performed on the differences between the fit values for models differing in one dimension. These results are reported in Table 4.

For participant 1, an MDS choice model using two dimensions did fit the responses better than an MDS choice model using only one dimension for day 2. By day 4, an MDS choice model using three dimensions did obtain a significantly higher fit value than an MDS choice model using only two dimensions. These results indicate that participant 1 used only one dimension on day 1, two dimensions on days 2 and 3, and three dimensions on days 4 and 5.[4] Similarly, the MDS analysis indicates that participants 2 and 3 used only one dimension on the first half of day 1, two dimensions on the second half of day 1, and three dimensions after day 1. Participant 4 appeared to use only one dimension on the first half of day 1, two dimensions on days 2 and 3, and three dimensions on days 4 through 6. Taken with the results from Experiment 1b, it appears that participants not only increased over days the number of cues used when categorizing, but also learned the correct (or relevant) cues to use to accurately categorize.

Given a task consisting of many dimensions, it is clear that participants begin by using only one dimension. Additional dimensions are then learned in a sequential fashion. What is remarkable from these data, is that participants learned to use all three dimensions. Dimension 1 had more variance than any of the other eight dimensions while dimension 2 had less variance than any of the other eight dimensions. Therefore, it is not surprising that participants were able to learn these two dimensions (i.e., the two dimensions out of nine that had differing variances). Dimension 3 on the other hand, had the same amount of variance as the six irrelevant dimensions, yet participants learned by the end of the experiment that this dimension was necessary for accurate categorization.

Conclusion

In conclusion, the studies reported here show that humans are able to learn artificial multidimensional categories. It was also shown that people are able to distinguish relevant from irrelevant dimensions in multidimensional categorization tasks. Results from such a task indicate that a satisficing model is best able to account for the participants' responses. In addition, the predictions made by the satisficing model regarding cue preference were shown to be in accord with the cue

protocol collected at the end of the experiment indicated that participants were only using three dimensions during Experiment 1a.

[4] Note, that on the last half of day 5, the increase in parameters used by and MDS choice model with three dimensions did not fit the data significantly better than an MDS choice model with less parameters (i.e., less dimensions).

Table 7: X^2_{diff} Values for Participants 1

	Participant 1		Participant 2		Participant 3		Participant 4	
Day/Half	MDS_1-MDS_2	MDS_2-MDS_3	MDS_1-MDS_2	MDS_2-MDS_3	MDS_1-MDS_2	MDS_2-MDS_3	MDS_1-MDS_2	MDS_2-MDS_3
1/1	8.34	0.08	3.26	0.3	8.62	3.6	1.02	2.84
1/2	6.56	6.76	27.18*	12.3	102.9*	18.84*	35.7*	5.08
2/1	83.3*	13.8	71.28*	18.96*	92.78*	9.94	86.16*	0.64
2/2	140.44*	2.56	69.94*	6.54	136.76*	30.16*	117.28*	3.62
3/1	214.98*	9.42	78.76*	22.04*	183.38*	29.14*	109.98*	0.38
3/2	174*	11.14	98.18*	35.86*	140.16*	21.1*	80.2*	4.8
4/1	244.36*	28.54*	116.86*	37.6*	155.02*	35.3*	74.56*	11.92*
4/2	146.22*	22.7*	149.28*	30.82*	196.44*	33.72*	80.36*	22.78*
5/1	151.78*	23.48*	116.8*	38.18*	113.6*	41.34*	80.48*	30.94*
5/2	201.98*	14.5	147.96*	34.34*	193.02*	39.92*	143.76*	18*
6/1	—	—	—	—	—	—	132.96*	37.92*
6/2	—	—	—	—	—	—	155.54*	33.08*

preferences of the participants. Finally, the new experimental design proposed provides a method for further testing the properties of dimensions (cues) that humans prefer (or are constrained?) to use.

References

Aeberhard, S., Coomans, D., & de Vel, O. (1994). Comparative analysis of statistical pattern recognition methods in high dimensional settings. Pattern Recognition, 27(8), 1065-1077.

Alfonso-Reese, L. A. (1995). General Recognition Theory toolbox for MATLAB. [Macintosh computer software], Santa Barbara, CA.

Ashby, F. G., & Gott, R. (1988). Decision rules in the perception and categorization of multidimensional stimuli. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 33-53.

Ashby, F. G., & Maddox, W. T. (1993). Comparing decision bound and exemplar models of categorization. Perception & Psychophysics, 53(1), 49-70.

Ashby, F. G., & Maddox, W. T. (1990). Integrating information from separable psychological dimensions. Journal of Experimental Psychology: Human Perception and Performance, 16(3), 598-612.

Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. Psychological Review, 95(1), 124-150.

Berretty, P. M., Todd, P. M., & Blythe, P. W. (1997). Categorization by elimination: A fast and frugal approach to categorization. In M. G. Shafto & P. Langley (Eds.), Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society. Mahwah, NJ: Erlbaum.

Berretty, P M, Todd, P M., and Martignon, L. (1999). Using few cues to choose: Fast and Frugal Categorization. In G. Gigerenzer & P M Todd (Eds.), Simple heuristics that make us smart. Oxford University Press.

Brainard, D. H. (1997). The Psychophysics Toolbox, Spatial Vision, 10, 443-446.

Maddox, W. T., & Ashby, F. G. (1998). Selective attention and the formation of linear decision boundaries: Comment on McKinley and Nosofsky (1996). Journal of Experimental Psychology: Human Perception and Performance, 24(1), 301-321.

Martin, G. L., & Pittman, J. A. (1991). Recognizing handprinted letters and digits using backpropagation learning. Neural Computation, 3(2), 258-267.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. Journal of Experimental Psychology: General, 115(1), 39-57.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Attention, similarity, and the identification-categorization relationship. Journal of Experimental Psychology: General, 115(1), 39-57.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies, Spatial Vision, 10, 437-442.

Posner, M. I.; Keele, S. W. (1968) On the genesis of abstract ideas. Journal of Experimental Psychology, 77(3), 353-363.

Quinlan, J. R. (1993). C4.5: Programs for machine learning. Los Altos: Morgan Kaufmann.

Tversky, A. (1972). Elimination by aspects: A theory of choice. Psychological Review, 79, 281-299.