Running Head: SEQUENCE EFFECTS

Sequence Effects in Categorization of Simple Perceptual Stimuli

Neil Stewart, Gordon D. A. Brown, and Nick Chater

University of Warwick

Stewart, N., Brown, G. D. A., & Chater, N. (2002). Sequence effects in categorization of simple perceptual stimuli. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28, 3−11.

Abstract

Categorization research typically assumes that the cognitive system has access to a (more or less noisy) representation of the absolute magnitudes of the properties of stimuli, and that this information is used in reaching a categorization decision. However, research on identification of simple perceptual stimuli suggests that people have poor representations of absolute magnitude information and that judgments about absolute magnitude are strongly influenced by preceding material. The experiments presented here show strong sequence effects in categorization tasks. Classification of a borderline stimulus was more accurate when preceded by a distant member of the opposite category than by a distant member of the same category. It is argued that this category contrast effect cannot be accounted for by extant exemplar or decision−bound models. The effect suggests the use of relative magnitude information in categorization. A memory and contrast model illustrates how relative magnitude information may be used in categorization.

Sequence Effects in Categorization of Simple Perceptual Stimuli

Categorization models are often divided into two general classes, each including a wide range of specific accounts: parametric Thurstonian decision−bound models (e.g., Ashby & Townsend, 1986), and non−parametric exemplar models (e.g., Medin & Schaffer, 1978; Nosofsky, 1986). However, all extant models of categorization assume that items can be represented in terms of their (more or less noisy) absolute location in a multidimensional space. This absolute location information is then assumed to be used in the decision process (either directly, as in exemplar models, or indirectly, in relation to decision bounds). Thus two key assumptions are (a) that the absolute location of a stimulus in multidimensional space, though possibly noisy, is available when a categorization or identification decision is made, and (b) that absolute location information provides the sole basis for categorization decisions. Formal models of categorization and identification based on these assumptions have a long history of successful application to a wide range of experimental paradigms (see Estes, 1994, for a review). Here we present evidence that relative magnitude information, derived from comparison of the current stimulus to recent stimuli, is also used in categorization.

Difficulty in Determining Absolute Magnitudes

Regarding the first assumption above, participants often have difficulty making accurate estimates of the absolute values of stimuli along simple perceptual dimensions, particularly in the absence of contextual information. For example, in a series of classic experiments by Garner (1954), participants' judgments of whether comparison tones were more or less than half as loud as a given reference tone were completely determined by the range of the comparison tones (see also Helson, 1964). Such a context effect should not be evident if participants did have access to absolute magnitude information. Baird, Green, and Luce (1980) demonstrated that two−thirds of the variability in loudness estimates was explained by the variability in the previous estimate when loudnesses were similar, suggesting the previous loudness is used as a reference point. Laming (1997) provided extensive discussion of these and other similar findings. These findings point to the importance of context and feedback in decision making and suggest that participants may have difficulty in accessing absolute magnitude information for these tasks. Of course, more or less accurate determination of absolute magnitude is often possible in other tasks; the information may be available directly or deduced from perception of the relative magnitude of the stimulus in comparison to an amalgam of reference or context stimuli. However, current models of classification assume absolute magnitude information is always available, and therefore these results suggest that such models may fail in situations where this information is unavailable.

It should be noted that the success of current models of categorization can not be taken as reason to ignore this problem: The use of random or controlled trial orders in almost all categorization experiments, followed by averaging over all stimuli of the same type, discards the very information about sequential context that may provide the true basis for categorization. Thus a primary aim of the research presented here is to examine sequence effects in categorization.

What Information is Used in Categorization?

The second key assumption embodied in many current categorization models is that categorization decisions are based on the (perceived or inferred) location of items in multidimensional space. It should be noted that this issue of information use can be examined separately from the related issue of information availability discussed above: even if accurate information about absolute magnitude is available, whether directly or indirectly, that information need not be used in identification and categorization decisions.

Much research demonstrates that the absolute identification of stimuli is heavily context dependent in that the response on trial n is influenced by the stimulus and response on trial $n-1$. In an absolute identification paradigm, participants are presented with stimuli that vary along a (normally uni−dimensional) psychological continuum (e.g., sounds that vary in amplitude, or lines of different lengths). Each stimulus is associated with a unique response. Normally the responses are arranged such that their order corresponds to the order of the stimuli in the psychological space. For example, if 10 line lengths are used − 1 cm, 2 cm, ... , and 10 cm − and the are 10 numbers for the responses − 1, 2, ... , and 10 − each line length would be associated with a single number. The 1−cm line could be associated with Response 1, the 2−cm line with Response 2, and so forth. On presentation of a stimulus, a participant is required to identify the unique response for that stimulus, before receiving the correct response as feedback. One crucial finding is that the response given to the current stimulus is assimilated to the immediately preceding stimulus (Garner, 1953; Holland & Lockhead, 1968; Hu, 1997; Lacouture, 1997; Lockhead, 1984; Luce, Nosofsky, Green, & Smith, 1982; Mori, 1989; Mori & Ward, 1995; Purks, Callahan, Braida, & Durlach, 1980; Staddon, King, & Lockhead, 1980; Ward & Lockhead, 1970, 1971). In other words, participants are systematically biased to respond as if the current stimulus is nearer the previous stimulus than it actually is. For example, if participants get Item 1 followed by Item 6, they will show a tendency to respond "5" instead of "6". The effect of stimuli further back in the sequence is the opposite and is referred to as a contrast effect (Holland & Lockhead, 1968; Lacouture, 1997; Ward & Lockhead, 1970, 1971). Thus, identification decisions depend on recent previous trials. Of course categorization decisions are not thought to be independent of previous trials, as it is precisely these trials that provide the information the categorization is based on. However, exemplar models do typically assume this information is not biased by the local sequential context provided by recent trials (for an account in terms of criterion shifting within a Thurstonian framework, see Luce et al., 1982; Treisman, 1985; Treisman & Williams, 1984).

Mori (1989) demonstrated that in absolute identification of uni−dimensional stimuli (e.g., frequencies or amplitudes), the information used by the decision process was limited to about 2.5 bits, and that this information was predicted almost completely by the current stimulus, the previous stimulus, and the previous response. The role of the previous stimulus and response in predicting the current response is further evidence that relative magnitude information is used in absolute identification. Consider the extreme case when the relation between recent successive trials and the current trial may solely determine the decision−making process. For example, in a binary categorization task an extreme possibility would be that each decision is made entirely on the basis of the perceived difference between the current and the previous stimulus, that is., with no reference to the absolute magnitude of the stimulus. It should be noted that this is a stronger claim than simply that decisions on successive trials are not independent. The claim is that the difference between stimulus on trial $n-1$ and the stimulus on trial n determines the response given to the stimulus on trial n. Participants would respond with the same category label as on the previous trial if there is a small difference between the two stimuli, and a different label if the difference is large. Indeed such a strategy would be the only one available to participants in the absence of absolute magnitude information. In more realistic situations, where partial absolute magnitude information is likely to be available, such a strategy is not likely to be used exclusively. However, for purposes of explication, we consider the extreme possibility that participants use only this memory and contrast (MAC) strategy. (A related concept is the "Bypass Rule"; Krueger & Shapiro (1981). Palmeri & Flanery (1999) also discussed the possibility that a similar strategy may be used to account for above−chance categorization in the absence of training.) The MAC strategy is far from being a general model of categorization. The model is used here only as an illustration of how relative magnitude information may be used in simple one−dimensional binary categorizations, and to investigate the predictions of the hypothesis that relative magnitude information is used in categorization.

MAC Strategy

Although intuition suggests that a MAC strategy will lead to very poor performance, preliminary modeling work indicates that strategies of this type can be surprisingly successful. Consider the case depicted in Figure 1 of 10 stimuli, equidistant from one another along a single dimension (such as loudness or pitch), and divided into two equal−sized categories. It is assumed that participants only have access to the magnitude and the direction of the difference between the current trial and the previous trial. It turns out that by optimizing the size of the difference needed to give a switch in categorization response, participants can achieve an accuracy of 85% in categorizing examples in a randomly ordered sequence of trials. (In fact, this observation is independent of the number of stimuli.) Such a model works by taking advantage of the correlation that exists between magnitude differences and category shifts when uni−dimensionally varying stimuli are involved. If a correct Category A response is given to Stimulus 1 on trial $n-1$, and there is a large positive dimensional shift up the scale to the stimulus on trial n, the large positive shift will be accompanied by a shift to a Category B response. A small shift, in contrast, is more likely to represent a within−category shift. An adaptive system could select the optimal shift size over which a change in responding should ensue. Although surprisingly successful, at least in the uni−dimensional case, this strategy will clearly lead to characteristic errors under particular circumstances. For example, if Item 1 is followed by Item 5 the large inter−trial difference will lead to an erroneous shift in response from Category A to Category B. In other words, large within−category shifts will induce errors. One can compare this to a large between−category shift, for example Stimulus 10 preceding Stimulus 5. The large shift will again cause a switch in response, this time correctly.

Traditional exemplar models make opposite predictions. Exemplar models can be adapted to predict sequence effects by assuming that more recent exemplars are more available in memory or weighted more heavily in the subsequent decision process (e.g., through the memory strength parameter of Nosofsky & Palmeri, 1997; see also Elliott &

Anderson, 1995). In exemplar models the probability of responding with a given category label is given by the ratio of the summed similarity to that category, divided by the summed similarity to all contending categories (i.e., in terms of Luce's, 1959, choice model). Therefore, the probability of responding with a given category can only be increased if exemplars of the same category are weighted more heavily in decision making, as when they have occurred very recently. The consequence of this is that when the item on the preceding trial is from the same category this must always lead to a greater tendency to respond with that category label, relative to the case where the previous stimulus was from the other category. This is the opposite prediction to that made by the MAC model described above.

Modeling

To support the intuitive argument above, categorization performance in a simple uni−dimensional random sequence was modeled using a MAC model and an exemplar model − the generalized context model (GCM; Nosofsky, 1986).

In this simple implementation of the MAC model, participants are assumed to base their categorization decision for the stimulus on trial n on the difference in response units, d, between the current stimulus and the stimulus on the preceding trial, trial $n-1$, and the experimenter feedback on trial $n-1$. Equation 1 uses Gaussian decay to relate the distance d to the probability of responding on trial n with the category label (i.e., feedback) from trial $n-1$.

$$P(\text{Same Category}) = e^{-cd^2} \tag{1}$$

The free parameter, c, determines the size of the distance required to give a change in category label by determining how quickly the probability of repeating the previous category label decreases as the difference between the previous and current stimuli increases. The Gaussian decay function was chosen because it is a smooth, monotonically decreasing function of d. Using Equation 1, the probability of a given response for the last stimulus in any pair of stimuli can be predicted. It should be noted that for some pairs (e.g., 5 followed by 1), the direction of the difference completely determines the categorization. There is no need to rely on the magnitude of the difference. For example, if it is known that Category A members take low values on the dimension, and the stimulus on trial $n-1$ is an A, any stimulus on trial n with a lower value, as indicated by the direction of the difference, must also be a member of Category A.

In a truly random sequence every pair of stimuli is equally likely. Therefore by calculating the probability of a correct response for every possible pair, and weighting all these probabilities equally, an average accuracy score can be obtained. The c parameter can then be fit to maximize accuracy. Figure 2 illustrates the predicted probabilities for each stimulus as a function of the preceding stimulus for the category structure illustrated in Figure 1. The predictions shown are for the optimal c parameter, which gave an accuracy of 85%. For the optimal c parameter the jump size that corresponded to an equal probability of responding with either category is 1.85 tones. However, overall accuracy remains very close to the maximum accuracy for a wide range of c parameters. Predictions for Stimuli 6−10 have been omitted, because, by the symmetry of the category structure, they are analogous to the predictions for Stimuli 5−1. Of interest is categorization accuracy of Stimulus 5, which is high when preceded by Stimulus 10, but low when preceded by Stimulus 1. In other words, a stimulus near the category boundary may be classified accurately when preceded by a distant member of the opposite category and poorly when preceded by a distant member of the same category. An exemplar model is unable to predict this pattern of results, as we show below.

Many, if not most, perceptual categorization experiments contain blocks where participants are not given trial−by−trial feedback. (In the experiments presented here participants were given trial by trial feedback.) It would be surprising if category contrast effects were not found in such conditions. Indeed, in absolute identification, in the absence of feedback, very similar sequence effects are observed, compared with those obtained with feedback (Ward & Lockhead, 1970, 1971). The MAC strategy described here assumes participants have knowledge of the correct categorization of previous stimuli. However, adaptation of the strategy to the no−feedback conditions is straightforward because of the correlation between the correct answer and the predicted answer. (Even in the simple MAC model presented here, where only information from trial $\underline{n}-1$ was used, accuracy was 85%.) A simple solution therefore would be to take the "correct" answer as that predicted by the model, that is, A if $\underline{P}(A) > .5$, otherwise B. Alternatively, the response on trial $\underline{n}$ could be a weighted mixture of the responses calculated for both possible categories of the stimulus on trial $\underline{n}-1$, i.e.,

$$P(A_n) = P(A_{n-1})\,e^{-\underline{c}\underline{d}^2} + [1 - P(A_{n-1})] + [1 - e^{-\underline{c}\underline{d}^2}] \;, \tag{2}$$

where $\underline{P}(A_{\underline{n}})$ is the probability of an A response on trial $\underline{n}$, $\underline{P}(A_{\underline{n}-1})$ is the probability of an A response on trial $n-1$, $\underline{d}$ is the difference between the stimulus on trial $\underline{n}$ and trial $\underline{n}-1$, and $\underline{c}$ is a free parameter determining the size of the distance required to give a change in category label, as in Equation 1.

The GCM is presented elsewhere (Nosofsky, 1986) but will be described briefly here. Each stimulus is represented by a vector in multidimensional space (i.e., the stimulus is represented using absolute magnitude information). Each stimulus encountered is stored, together with its category label. The probability with which a Stimulus $\vec{x}$ is classified into category $\underline{C}_k$, $\underline{P}(\underline{C}_k|\vec{x})$, is given by the ratio of its summed similarity to examples of that category, $\underline{h}_k(\vec{x})$, divided by the summed similarity to all contending categories:

$$P(\underline{C}_k|\vec{x}) = \frac{\beta_k \underline{h}_k(\vec{x})}{\sum_{i=1}^{K} \beta_i \underline{h}_i(\vec{x})} \tag{3}$$

where $\underline{\beta}_i$ is the bias to respond with category $\underline{i}$. Similarity is a monotonically decreasing function of distance, and is typically either an exponential decay or a Gaussian. Thus,

$$\underline{h}_k(\vec{x}) = \sum_{i=1}^{N_k} e^{-\underline{c}\underline{d}(\vec{x}, \vec{x}_i)^q} \tag{4}$$

where $d(x, x_i)$ is the distance between Stimulus x and Stimulus x_i in psychological space, q specifies the form of the similarity function, and c is a free parameter for the discriminability of the stimuli.

The GCM can be adapted to predict sequence effects by weighting the stimulus on the previous trial more heavily in the summed similarity calculations. In intuitive terms, this corresponds to the stimulus either being more available in memory or being weighted more heavily in the decision process. This means the current stimulus will always be more similar to the category of the preceding stimulus than it would be with no weighting. To demonstrate clear sequence effects, the stimulus on the previous trial was arbitrarily weighted 10 times more heavily than other stimuli. The GCM was used to predict classification accuracies for the category structure described in Figure 1. Figure 3 shows the categorization accuracy for the stimulus on trial n as a function of the preceding stimulus on trial $n-1$ for the GCM (q=2, c=0.25, no category bias). Although the exact predictions depend on the generalization parameter, c, and the choice of similarity function parameter, q, the qualitative pattern of results is independent of these choices. The optimal value of the c parameter is infinite, as then there will be no generalization between stimuli, and performance will be 100% accurate, with no effect of the previous stimulus. The size of the weighting for the stimulus on trial $n-1$ also does not affect the qualitative pattern − a larger weighting simply makes the pattern more extreme. The GCM, unlike the MAC model, is always constrained to predict more accurate classification in the case when the preceding stimulus is from the same category rather than the opposite category.

Overview of Experiments

In the experiments in this article, these opposing predictions for the relative accuracy of classification of a borderline stimulus, preceded by either a distant member of the same category or a distant member of the other category, are tested. Both of the experiments use the category structure in Figure 1. The aim was to demonstrate a category contrast effect, whereby classification of borderline stimuli is more accurate when preceded by a distant stimulus from the other category than by a distant stimulus from the same category. A MAC strategy would be able to offer an account of this intuitive potential result, but existing models of categorization would not. The existence of a category contrast effect would therefore provide evidence that categorization is based, at least in part, on relative location information. Experiment 1 uses the frequency of a tone as the dimension of variability in a simple binary classification. Experiment 2 uses simple geometric figures used in categorization experiments where participants have typically been hypothesized to categorize on the basis of absolute magnitude information alone.

Experiment 1

Experiment 1 aims to demonstrate a simple category contrast effect using the category structure in Figure 1. As the concern is with the effect of distant stimuli on the classification of stimuli on the borderline between the two categories, these pairs of stimuli (1 before 5, 10 before 5, 1 before 6, and 10 before 6) were overrepresented in pseudorandom sequences, so that enough data could be gathered in a short experiment. The pseudo random sequences are controlled so that the runs of consecutive categorization responses, the relative frequencies of each tone, and the relative frequencies of each sized jump between tones would be as found in a truly random sequence.

Method

Participants. Ten University of Warwick undergraduates participated in this 10−min experiment.

Stimuli. Ten 500−ms sine−wave tones of differing frequency were used as stimuli in this experiment. Each tone was 1% higher in frequency than the tone immediately lower in frequency, and thus the tones were equally spaced on a log−frequency scale. The first tone had a frequency of 600.00 Hz, and the last tone had a frequency of 656.21 Hz. The intention was that adjacent tones anywhere along the scale would be equally discriminable when presented in isolation.

Design. The 10 tones were divided into two categories, with the 5 lowest frequency tones in one category, and the 5 highest frequency tones in the other category. Tones were presented sequentially for categorization. Of interest in this experiment are the effects of the immediately preceding tone (trial $n-1$) on the categorization of the current tone (trial n). Numbering the tones from 1 (lowest frequency) to 10 (highest frequency), the four critical pairs of tones are 1 before 5, 10 before 5, 10 before 6 and 1 before 6. The pairs 1→5 and 10→6 contain a tone distant in frequency space followed by a borderline member of the same category. The pairs 10→5 and 1→6 contain a distant tone followed by a borderline member of the other category. A simple comparison of the proportion correct on the last trial of each pair for the two pair types (either within category or between category) will allow exemplar and MAC accounts to be distinguished.

Each critical pair was presented once in each block of 20 trials. The four critical pairs were assigned at random to the 4th and 5th, 9th and 10th, 14th and 15th, and 19th and 20th trials in a block. The remaining tones − 2, 3, 4, 7, 8 and 9 − were placed in the unfilled trials at random, subject to the following constraints: (a) each tone occurs equally frequently, (b) the number of occurrences of each size jump in frequency reflects the natural distribution of these jumps for a random stream of 10 tones, (c) the lengths of runs of tones of the same category is fixed to mimic a random sequence. With these constraints, only 42 possible sequences can be generated. For each block a sequence was selected at random from one of the possible sequences. The constraints were designed to allow the critical pairs to be over−represented in a sequence without the sequence seeming nonrandom.

Procedure. Participants were tested one at a time in a quiet room. Participants were instructed that they would hear a number of tones, one after the other. They were told that after each tone they would be asked to respond with one of two labeled keys depending on which category they thought the tone came from. Participants were asked to respond as quickly as possible without making mistakes. Although at first participants would have to guess, they were informed that by attending to the correct answer displayed on the screen after each response, they could learn which tones belonged to which category. They were given an opportunity to ask the experimenter questions before the experiment began.

Ten blocks of 20 trials were presented to each participant. For each block a different pseudorandom sequence, as described in the design, was randomly chosen. Each trial began with a tone, presented for 500 ms, over Sony DR−S3 closed−back headphones. Tones were generated by, and responses were gathered, using an Apple Macintosh Performa 475 computer. A "?" prompt appeared on the screen with the onset of the tone. From the onset of the tone participants were able to respond with either Z or X (labeled "A" and "B" respectively) on a standard keyboard. The assignment of labels to categories was counterbalanced across participants. The "?" prompt disappeared immediately after participants responded. After the participants had responded, or 1,500 ms after the offset of the tone, whichever was later, the correct answer was displayed on the screen for 1,000 ms. There was a 500−ms pause before the next trial began. Participants completed all blocks with no breaks between blocks. The experiment took about 10 min to complete.

Results

Categorization accuracy reached an asymptote of about 90% correct after the first block of 20 trials. Performance on the last tone in a critical pair is shown as a function of whether the first tone of the pair came from the same category or the other category (Figure 4). There was a large difference in performance in the two pair types, with participants classifying a borderline tone significantly more accurately after a distant tone from the other category, compared to a distant tone from the same category, $t(9)=3.67$, $p<.01$. This pattern is the same for both pairs 1→5 and 10→5, and for pairs 10→6 and 1→6, and is consistent with a MAC hypothesis.

An alternate explanation of these results needs to be ruled out. The difference between the two types of critical pairs is that to get both tones in a within−category pair correct participants must make the same category response twice in a row, but to get both tones correct in a between category pair correct participants must switch responses. Thus, if participants are biased against making two identical responses in a row, participants would show poorer accuracy on the final tone of the within−category pair than on the final tone of the between−category pair. To eliminate this possibility, responses to filler items were examined to measure possible bias. Participants were only slightly more likely to persevere with a response than they should have been, given the sequence they were presented with. This deviation was not significant, $t(9)=0.47$, $p=.65$, and is in the wrong direction to explain the pattern of responding on the last item in the critical stimuli pairs.

Discussion

In categorizing a sequence of tones, categorization decisions are influenced by the immediately preceding tone. This finding is consistent with evidence from absolute identification, where there are also strong sequence effects (Lacouture, 1997; Mori, 1989; Mori & Ward, 1995; Ward & Lockhead, 1970, 1971). When categorizing a tone on the borderline between the two categories, a preceding, large within−category shift induced significantly more errors than a between−category shift (i.e., a category contrast effect is demonstrated). This effect is consistent with a MAC strategy but not with an exemplar−based strategy. Although these effects could potentially be explained by a simple alternation bias (Dember & Richman, 1985), analysis of filler trials in the pilot experiment reveals no evidence of such bias. The category contrast effect is strong evidence that participants' categorizations are based on the relative frequency of the current and preceding tones (possibly in addition to absolute magnitude information).

It is possible that exemplar models may be able to predict successfully the category contrast effect observed when sequence effects in absolute magnitude estimation are taken into consideration. The modeling using the exemplar model in the introduction did not take identification assimilation or contrast effects into account. Consideration is given here to the predictions of an exemplar model when sequence effects in identification are used as a potential explanation of the sequence effect in categorization demonstrated in Experiment 1. In an absolute identification task, the response given to the current stimulus is assimilated to the immediately preceding stimulus (Garner, 1953; Holland & Lockhead, 1968; Hu, 1997; Lacouture, 1997; Lockhead, 1984; Luce et al., 1982; Purks et al., 1980; Staddon et al., 1980; Ward & Lockhead, 1970, 1971). How would such assimilation affect an exemplar model's predictions for the critical pairs of interest? When the distant tone is from the same category, assimilation should cause participants to perceive the tone as more similar to the exemplars of the correct category, and less similar to the exemplars of the incorrect category, than it really is. Identification assimilation will therefore increase categorization accuracy when the preceding tone is from the same category. Assimilation when the distant tone is from the other category will cause participants to perceive the current tone as more similar to the category of the preceding tone, and therefore more similar to the other category, and thus participants will be more likely to categorize it incorrectly. Therefore, identification assimilation could cause the exemplar model to predict participants to be even more likely to be correct on a borderline tone when it is preceded by a distant member of the same category and even less accurate when it is preceded by a distant member of the other category. This effect is in the opposite direction than that needed to allow an exemplar model to explain the pattern of performance observed in Experiment 1. However, if with identification of frequency there is an identification contrast effect, then an exemplar model would be able to account for the results. There is no evidence for identification contrast to the immediately preceding item in absolute identification of frequency, and indeed such an effect would be at odds with the assimilation observed for other dimensions in previous research. However, such a bias in identification could explain the category contrast effect, without assuming a MAC strategy, as follows. For the within−category pair, participants would perceive the borderline tone as further from the distant tone, making it less similar to the correct category than it really is and more similar to the incorrect category. For the between−category pairs participants would perceive the borderline tone as more similar to the correct category than it really should be and less similar to the incorrect category. Thus the exemplar model could predict performance to be higher for the different category critical pairs than for the same category pairs, as observed in Experiment 1. An experiment from our laboratory (Stewart, 2001) has demonstrated within−participants assimilation in identification and contrast in categorization in alternate blocks for tones varying in frequency. Thus, as well as replicating the basic category contrast effect the experiment shows that the sequence effects in identification do not allow prediction of contrast in categorization.

Experiment 2

Experiment 2 is very similar to Experiment 1. The main difference is that the tones were replaced with simple visual stimuli. The stimuli are those used by Nosofsky (1985, 1986) − semicircles that vary in radius, with radial lines that vary in orientation. These stimuli were selected because they are typical of stimuli used in categorization experiments (Ashby & Gott, 1988; Ashby & Waldron, 1999; Maddox & Ashby, 1993; Nosofsky, 1985, 1986). Models of categorization applied to data from research with such stimuli assume participants represent the stimuli in a multidimensional space, and therefore make the implicit assumption that participants have access to absolute magnitude information for these stimuli (e.g., the GCM, Nosofsky, 1986; and general recognition theory or decision−bound theory, Ashby & Townsend, 1986). If a contrast effect can be demonstrated with these stimuli, then this would demonstrate the generality of the category contrast effect.

Although for the purposes of explication of the MAC strategy we have been assuming that participants do not have absolute magnitude information available to them, we certainly do not claim that this information is completely unavailable. With the simple visual stimuli used in this experiment, it is possible that participants have some absolute magnitude information, either directly, or by comparison with the context (e.g., the edges of the monitor). Whether this information is directly perceived or deduced from the context the stimuli are presented in is not at issue. However the fact that that the information may be available means that it may be used to inform categorization decisions. If this is the case, the category contrast effect is expected to be smaller. Accordingly more participants were tested in Experiment 2 than in Experiment 1 to detect a potentially smaller effect.

Method

Participants. Twenty−six University of Warwick undergraduates and postgraduates participated.

Stimuli. The stimuli used in this experiment were semicircles of varying radius, with radii of varying angle, as used by Nosofsky (1985, 1986). In Nosofsky’s (1986) experiment, four possible semicircle radii were crossed with four possible radius orientations, to create 16 possible stimuli, arranged in a 4 x 4 grid in diameter−orientation space. In this experiment, 10 different stimuli were created, arranged in a straight line in diameter−orientation space. Thus, both semicircle radius and radius orientation were diagnostic of category. Two alternative spacings of the 10 stimuli were considered. Although 10 stimuli spaced equally across a diagonal of Nosofsky’s (1986) square of stimuli would equate the overall area of stimulus space used by the stimuli, this solution was rejected because the 10 stimuli would be less discriminable from one another than Nosofsky’s (1986) stimuli, as they fill the stimulus space more densely. It was felt that this would hinder participants in the possible application of an exemplar strategy, as the stimuli would be more confusable. An alternative arrangement (Figure 5) where the 10 stimuli extend outside the region of space occupied by Nosofsky’s (1986) stimuli was used. Each adjacent pair of stimuli was then spaced as in Nosofsky’s experiment. It should be noted that this choice of stimuli is the conservative choice, favoring exemplar models. Stimuli were presented for only 150 ms, as in Nosofsky’s (1986) original experiment. Although stimuli do vary on two dimensions, the two dimensions are perfectly correlated. Thus, we cannot exclude the possibility that participants may treat it as a uni−dimensional task.

Design and procedure. The design and procedure are the same as in Experiment 1, except tones were replaced with a 150−ms presentation of a semicircle with line stimulus in green pixels on a black background.

Results

This analysis is identical to that performed for Experiment 1. As in Experiment, participants quickly reached asymptotic performance of over 90% of filler stimuli correct after one block of trials. Two participants were eliminated from the study for spontaneously reporting that they realized that certain pairs were designed to trick them, and responding to counter this effect. A further participant was eliminated for failing to perform above chance on filler items throughout the experiment. (It should be noted that the filler items were categorized almost perfectly by all other participants.) For the remaining participants, performance on the last semicircle in a critical pair is shown as a function of whether the first semicircle of the pair came from the same category or the other category (Figure 6). There was a smaller difference in performance between the two pair types than in Experiment 1. The difference, however, was significant, with participants classifying a borderline stimulus significantly more accurately after a distant semicircle from the other category, compared with a distant semicircle from the same category, $t(22)=3.66$, $p<.01$. This pattern is the same for both pairs 1→5 and 10→5, and for pairs 10→6 and 1→6, and is consistent with the MAC strategy.

As in Experiment 1, responses to filler items were examined to measure possible bias. Participants were very slightly more likely to persevere with a response than they should be. This difference was not significant, $t(22)=1.39$, $p=.17$, and such a perseverance bias could not explain participants' worse performance in the same condition. (A bias towards giving the same response would reduce errors for these critical pairs, and increase errors for the different category critical pairs.)

Discussion

The category contrast effect demonstrated in Experiment 1 was replicated in Experiment 2 using different stimuli. The effect was approximately half the size of the effect observed in Experiment 1, consistent with the hypothesis that participants have increased access to absolute magnitude information (but see the General Discussion for an alternative explanation consistent with the MAC account). However, the effect is still large, and constitutes a demonstration of a sequence effect in categorization that cannot be accounted for by models that assume that categorization is based only on absolute magnitude information.

General Discussion

A category contrast effect has been demonstrated whereby categorization accuracy of a stimulus near the boundary between two categories is higher when preceded by a distant stimulus from the opposite category than by a distant stimulus from the same category. This large effect persisted throughout each experiment, even after average accuracy reached over 85%. Experiment 1 demonstrated this effect in a binary classification of tones varying in frequency. Experiment 2 replicated this effect using simple visual stimuli, extending the generality of the result. We also found the effect in a meta−analysis of data from other categorization experiments where similar, simple geometric figures were used as stimuli (Stewart, 2001). In these experiments random trial ordering was used, and thus the category contrast effect is not an artifact of the pseudorandom sequences used here. Although the category contrast effects could potentially be explained by a simple response alternation bias (Dember & Richman, 1985), analysis of the bias in filler trials in these experiments discounts this explanation.

The existence of this category contrast effect provides two challenges to existing models of categorization. First, the assumption that categorization is based only on absolute magnitude information is challenged, by the demonstraton of a pattern of errors that can only be accounted for by participants' (at least partial) reliance on relative magnitude information. Second, the category contrast effect provides evidence that the local sequential context biases categorization decisions. The category contrast effect is consistent with the MAC account presented here, where classification of a stimulus is based on comparison with the preceding stimulus.

The modeling presented demonstrates that current exemplar models cannot easily account for the results. Decision−bound models are unable to account for the results when they are adapted to assume the location of the decision bound is altered by preceding material. To model the category contrast effect, the decision bound would need to move towards the preceding stimulus, so that a borderline stimulus from the same category would fall inside the other category (see, e.g., Treisman, 1985; Treisman & Williams, 1984). This movement of the decision bound would also therefore predict contrast effects in identification, inconsistent with the observed assimilation effects. At present, we offer no account of why contrast effects are observed in categorization but assimilation effects are observed in absolute identification.

Other Sequence Effects in Categorization

Other researchers have investigated sequence effects in categorization. Medin and Bettger (1994) demonstrated that the sequence of training exemplars altered later recognition performance − when training exemplars were sequenced to maximize similarity between adjacent items old/new recognition was improved. Elliott and Anderson (1995) manipulated the order of presentation of training exemplars, showing that more distant items were less available for use in a categorization decision, with the decay following a power law. The number of intervening items was shown to be more important than the intervening time. The concern of these researchers was with the longer term effects of the sequence manipulations and not with the local sequence effects investigated here.

The Magnitude of the Category Contrast Effect

The sizes of the category contrast effects demonstrated here are smaller than those predicted by the simple MAC strategy. There are two potential reasons for this: First, participants may be using an improved MAC strategy where comparisons with tones further back in the sequence also inform the categorization decision. Use of this additional information would improve classification accuracy, therefore reducing the size of the category contrast effect. In fact formal modeling has shown inclusion of information from trial $n-2$ divides the size of the effect in half (if the Stimulus $n-1$ to n difference is weighted equally with the Stimulus $n-2$ to n difference − though this is not the optimal way to combine these two sources of information). Thus the smaller category contrast effect observed in Experiment 2 may be explained if participants are better able to use information from preceding trials when stimuli are simple geometric figures rather than tones.

The second possibility is that participants have partial access to absolute magnitude information, and make use of this information. In making an absolute identification of length decision, reducing the luminance of lines (with the intention of reducing the amount of absolute magnitude information available) increased the relative contribution information from previous trials compared with information from the current stimulus (Mori, 1989). The idea that reduced availability of absolute magnitude information increases the reliance on a MAC−like strategy is consistent with the pattern of the size of effects observed in the experiments presented here. The category contrast effect was larger in Experiment 1, where tones varying in frequency were used, than in Experiment 2, where simple visual stimuli were used. This is consistent with the assumption that these simple visual stimuli allow participants more access to absolute magnitude information (either perceived directly, or deduced from comparison with the presentation context).

Conclusions

A sequence effect in categorization has been demonstrated that challenges the assumption, implicit in existing models of categorization, that categorization is based <u>only</u> on absolute magnitude information. An alternate model has been presented that accounts for this effect by assuming participants instead rely on comparison of a stimulus to immediately preceding stimuli to make a categorization decision.

References

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. Journal of Experimental Psychology: Animal Behavior Processes, 14, 33−53.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. Psychological Review, 93, 154−179.

Ashby, F. G., & Waldron, E. M. (1999). On the nature of implicit categorization. Psychonomic Bulletin & Review, 6, 363−378.

Baird, J. C., Green, D. M., & Luce, R. D. (1980). Variability and sequential effects in cross−modality matching of area and loudness. Journal of Experimental Psychology: Human Perception and Performance, 6, 277−289.

Dember, W. N., & Richman, C. L. (1985). Spontaneous alternation behavior. New York: Springer−Verlag.

Elliott, S. W., & Anderson, J. R. (1995). Effect of memory decay on predictions from changing categories. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 815−836.

Estes, W. K. (1994). Classification and cognition. New York: Oxford University Press.

Garner, W. R. (1953). An informational analysis of absolute judgments of loudness. Journal of Experimental Psychology, 46, 373−380.

Garner, W. R. (1954). Context effects and the validity of loudness scales. Journal of Experimental Psychology, 48, 218−224.

Helson, H. (1964). Adaptation−level theory. New York: Harper & Row.

Holland, M. K., & Lockhead, G. R. (1968). Sequential effects in absolute judgments of loudness. Perception and Psychophysics, 3, 409−414.

Hu, G. (1997). Why is it difficult to learn absolute judgment tasks? Perceptual and

Motor Skills, 84, 323−335.

Krueger, L. E., & Shapiro, R. G. (1981). Inter−trial effects of same−different judgements. Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 33, 241−265.

Lacouture, Y. (1997). Bow, range, and sequential effects in absolute identification: A response−time analysis. Psychological Research, 60, 121−133.

Laming, D. (1997). The measurement of sensation. London: Oxford University Press.

Lockhead, G. R. (1984). Sequential predictors of choice in psychophysical tasks. In S. Kornblum & J. Requin (Eds.), Preparatory states and processes (pp. 27−47). Hillsdale, NJ: Erlbaum.

Luce, R. D. (1959). Individual choice behavior. New York: Wiley.

Luce, R. D., Nosofsky, R. M., Green, D. M., & Smith, A. F. (1982). The bow and sequential effects in absolute identification. Perception & Psychophysics, 32, 397−408.

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

Medin, D. L., & Bettger, J. G. (1994). Presentation order and recognition of categorically related examples. Psychonomic Bulletin & Review, 1, 250−254.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. Psychological Review, 85, 207−238.

Mori, S. (1989). A limited−capacity response process in absolute identification. Perception & Psychophysics, 46, 167−173.

Mori, S., & Ward, L. M. (1995). Pure feedback effects in absolute identification. Perception & Psychophysics, 57, 1065−1079.

Nosofsky, R. M. (1985). Overall similarity and the identification of separable−dimension stimuli: A choice model analysis. Perception & Psychophysics, 38, 415−432.

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

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar−based random walk model of speeded classification. Psychological Review, 104, 266−300.

Palmeri, T. J., & Flanery, M. A. (1999). Learning about categories in the absence of training: Profound amnesia and the relationship between perceptual categorization and recognition memory. Psychological Science, 10, 526−530.

Purks, S. R., Callahan, D. J., Braida, L. D., & Durlach, N. I. (1980). Intensity perception. X. Effect of preceding stimulus on identification performance. Journal of the Acoustical Society of America, 67, 634−637.

Staddon, J. E. R., King, M., & Lockhead, G. R. (1980). On sequential effects in absolute judgment experiments. Journal of Experimental Psychology: Human Perception & Performance, 6, 290−301.

Stewart, N. (2001). Perceptual categorization. Unpublished doctoral dissertation, University of Warwick, England.

Treisman, M. (1985). The magical number seven and some other features of category scaling: Properties for a model of absolute judgment. Journal of Mathematical Psychology, 29, 175−230.

Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. Psychological Review, 91, 68−111.

Ward, L. M., & Lockhead, G. R. (1970). Sequential effect and memory in category judgment. Journal of Experimental Psychology, 84, 27−34.

Ward, L. M., & Lockhead, G. R. (1971). Response system processes in absolute judgment. Perception & Psychophysics, 9, 73−78.

Author Note

Neil Stewart, Gordon D. A. Brown, and Nick Chater, Department of Psychology, University of Warwick.

The research was partly funded by a Graduate Assistantship awarded to the first author by the University of Warwick, a Leverhulme Trust Grant (#F/215/AY) awarded to the Gordon D. A. Brown and Nick Chater, and a Biotechnology and Biological Sciences Research Council grant 88/S09589 awarded to Evan Heit. Nick Chater was supported by European Commission Grant RTN−HPRN−CT−1999−00065. We thank Lewis Bott, Evan Heit and Koen Lamberts for their helpful comments. We also thank Suzanna Bootle for running Experiment 1.

Correspondence concerning this article should be addressed to Neil Stewart, Department of Psychology, University of Warwick, Coventry, CV4 7AL, United Kingdom. E−mail: neil.stewart@warwick.ac.uk.

Figure Captions

Figure 1. Ten stimuli distributed evenly along a single psychological dimension divided into two categories.

Figure 2. The predictions for the memory and contrast model for the simple category structure illustrated in Figure 1. Accuracy for a stimulus on trial n is plotted as a function of the stimulus on trial $n-1$.

Figure 3. The predictions for the generalized context model for the simple category structure illustrated in Figure 1. Accuracy for a stimulus on trial n is plotted as a function of the stimulus on trial $n-1$.

Figure 4. The proportion of correct responses for same category tone pairs (1→5 and 10→6) and different category pairs (1→6 and 10→5) for Experiment 1. Error bars represent the standard errors of the mean.

Figure 5. The stimulus structure used in Experiment 2 compared with Nosofsky's (1986) stimulus structure.

Figure 6. The proportion of correct responses for same category stimulus pairs (1→5 and 10→6) and different category pairs (1→6 and 10→5) for Experiment 2. Error bars represent the standard errors of the mean.

Figure 1

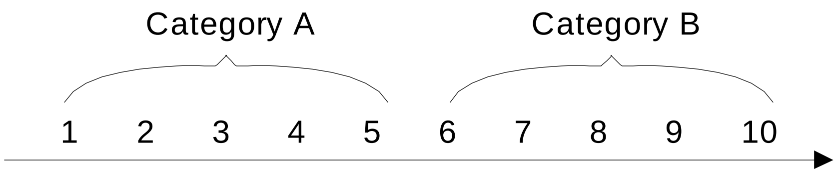


Figure 2

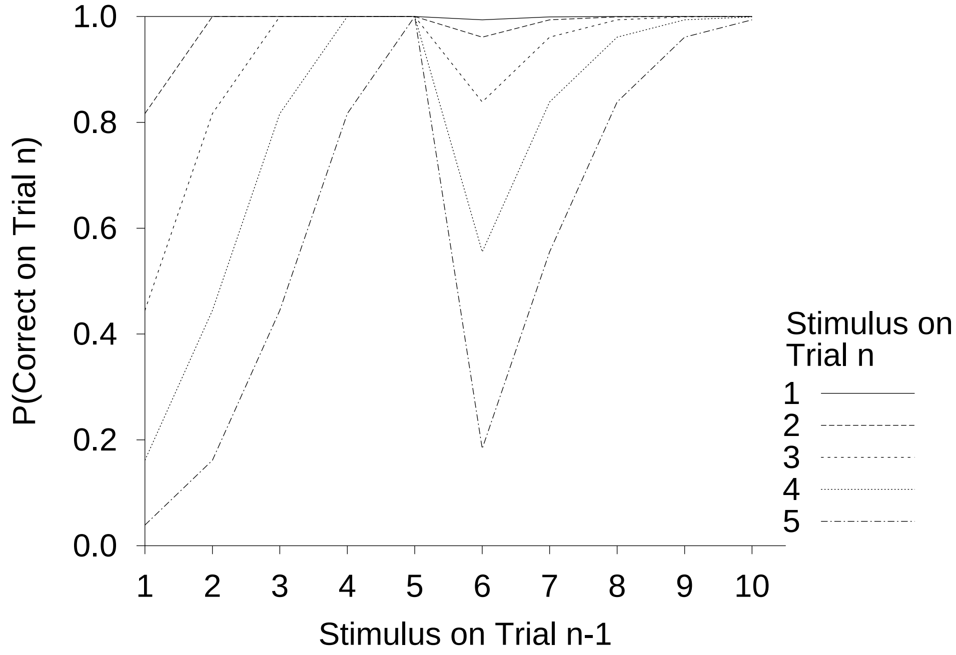

1.0
0.8
0.6
0.4
0.2
0.0
P(Correct on Trial n)
1
2
3
4
5
6
7
8
9
10
Stimulus on Trial n-1
Stimulus on Trial n
1
2
3
4
5

Figure 3

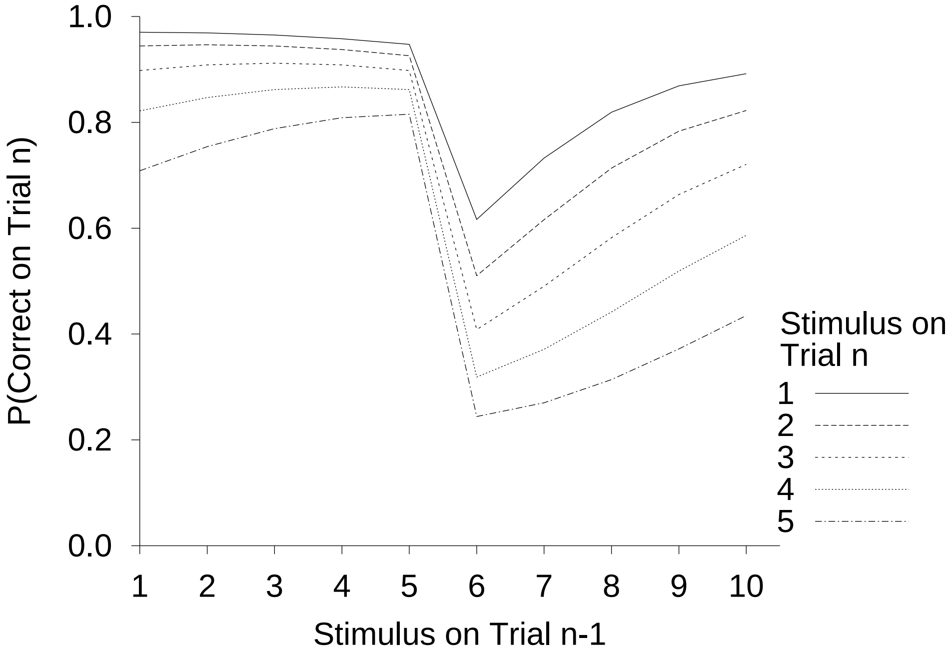

1.0
0.8
0.6
0.4
0.2
0.0
P(Correct on Trial n)
1
2
3
4
5
6
7
8
9
10
Stimulus on Trial n-1
Stimulus on Trial n
1
2
3
4
5

Figure 4

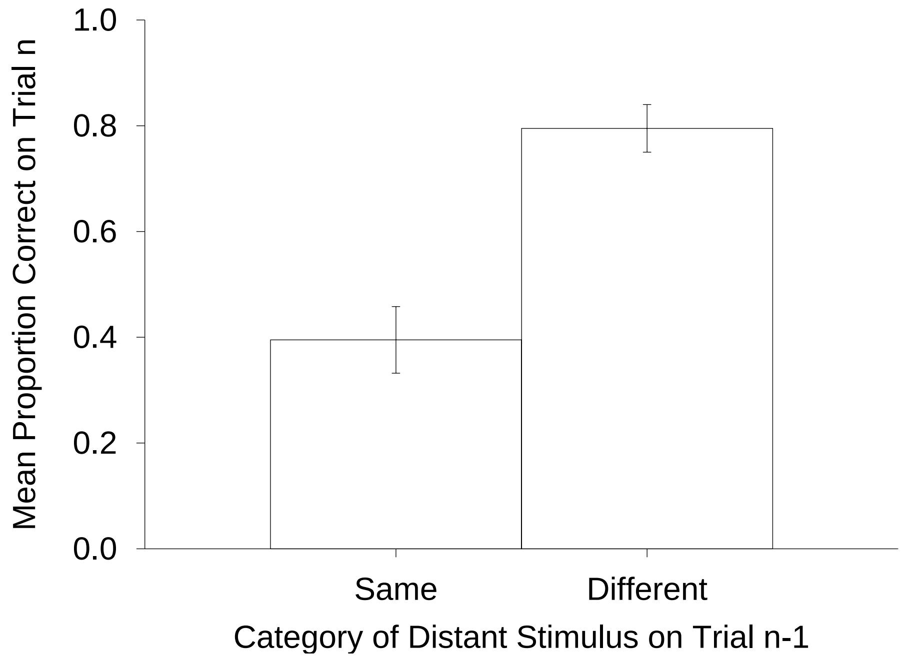

Mean Proportion Correct on Trial n
1.0
0.8
0.6
0.4
0.2
0.0
Same
Different
Category of Distant Stimulus on Trial n-1

Figure 5

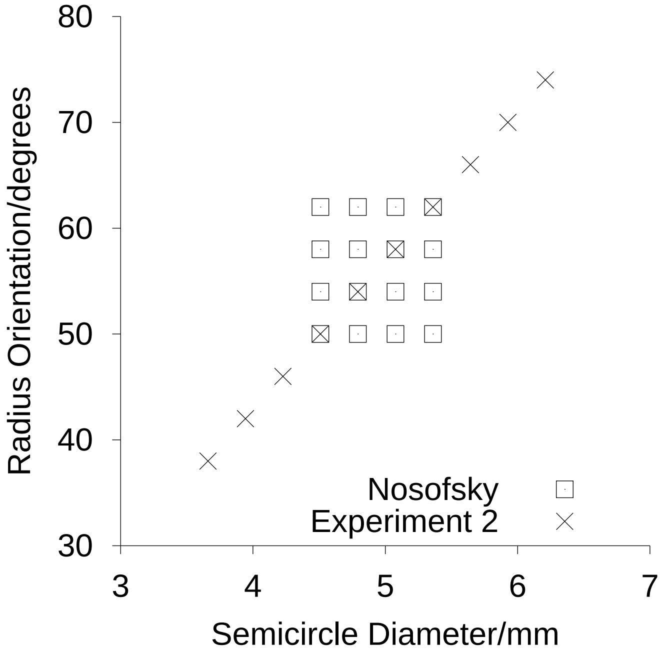
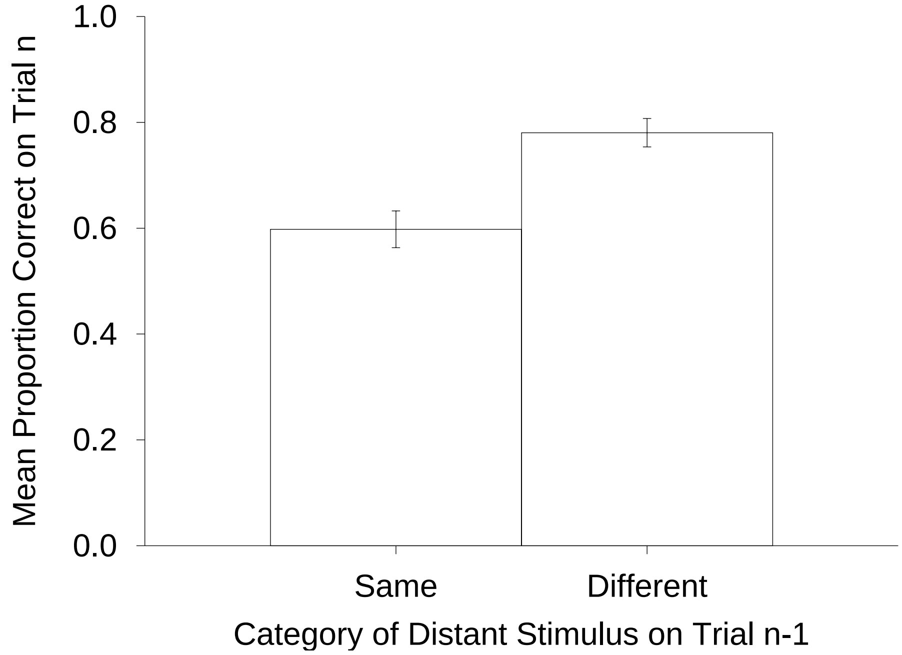

80
70
60
50
40
30
3
4
5
6
7
Radius Orientation/degrees
Semicircle Diameter/mm
Nosofsky
Experiment 2

Figure 6


1.0
0.8
0.6
0.4
0.2
0.0
Mean Proportion Correct on Trial n
Same
Different
Category of Distant Stimulus on Trial n-1