Journal of Experimental Psychology:
Human Perception and Performance
1996, Vol. 22, No. 3, 634–661

0096-1523/96/$3.00

A Constructive–Associative Model of the Contextual Dependence of Unidimensional Similarity

Douglas H. Wedell
University of South Carolina

Conditions under which pairwise dissimilarity ratings should reflect manipulations of the stimulus distribution were outlined by a model that proposed these effects. These conditions arise from either a context dependent process for constructing implicit scale values or a process that uses previously established stimulus–response associates. Consistent with the model, results from 3 experiments using unidimensional psychophysical stimuli demonstrated disordinal context effects on pairwise dissimilarity ratings when (a) there was a 3-s delay between presentation of pair members or (b) a unidimensional rating task preceded the pairwise dissimilarity ratings. Global effects of density were fit well by a model that extended A. Parducci's (1983) range–frequency theory to dissimilarity ratings. Local density effects were generally consistent with predictions from C. L. Krumhansl's (1978) distance–density theory.

The contextual relativity of judgment is well established and has been extensively researched within psychology for nearly a century (e.g., Beebe-Center, 1929; Helson, 1947; Hollingsworth, 1910; Hunt & Volkmann, 1937; Johnson, 1944). The same stimulus may be judged high or low on a dimension, depending on the values of the other contextual stimuli that are experienced. For example, a person who is judged tall when standing among mostly shorter persons may be judged short when standing among mostly taller persons. Although this type of contextual contrast is common, there is controversy surrounding its psychological status. This dispute can be framed using the height example as follows: Does the cognitive representation of the person's height differ between contexts, or does the judge simply choose different categorical responses in the different contexts because of response constraints?

If contextual effects occur at the representational level, then they would be expected to have pervasive effects across a variety of tasks. Tversky (1977), in challenging traditional geometric models of similarity, argued that some contextual manipulations do alter the psychological representation of the similarity structure of stimuli. Using schematic faces as stimuli, he demonstrated that which of two faces was judged more similar to a target face depended on the features of a third face included in the judgment set. Tversky (1977) interpreted these effects in terms of a feature-based model of similarity and argued that these and other contextual effects on similarity measures violate the basic axioms of traditional geometric models. Subsequent research has provided substantial support for the feature-based model of similarity (Gati & Tversky, 1984; Tversky & Gati, 1982).

In response to criticisms of geometric models of similarity, Krumhansl (1978) presented a distance–density model that could account for many of the contextual effects described by Tversky (1977). A critical property of the Krumhansl model was that similarity depended not only on the interpoint distance between two stimuli in a psychological space, but also on the relative density of the contextual stimuli in the space. Similarity was proposed to be inversely related to density so that the same interpoint distance would correspond to greater similarity when the stimuli were located in a sparse rather than a dense region of the psychological space.

This inverse relationship between similarity and stimulus density was derived from and consistent with contextual effects of density manipulations on unidimensional ratings, as described by Parducci's (1963) range–frequency theory of judgment. Specifically, the frequency principle of that theory states that the judged value of the stimulus is proportional to its rank in the set of contextual stimuli (Parducci, 1983). This theoretical principle implies that the form of the mean rating function should mimic the form of the (percentile) rank function, an implication that has been verified in numerous experiments using both psychophysical and social judgments (e.g., Birnbaum, 1974; Mellers, 1983; Mellers & Birnbaum, 1983; Parducci & Perrett, 1971; Parducci & Wedell, 1986; Riskey, Parducci, & Beauchamp, 1979; Smith, Diener, & Wedell, 1989; Wedell, 1994; Wedell & Parducci, 1988; Wedell, Parducci, & Roman, 1989).

As an example of how the frequency principle operates, consider the typical situation in which the contextual distribution is operationally defined as the entire set of stimuli being judged in the experiment. One group of participants may be exposed to a positively skewed distribution, with stimulus values 1, 2, 3, 4, and 5 being presented with

This work was supported by Biomedical Research Support Grant 2-SO7-RR07160 at the University of South Carolina.

Correspondence concerning this article should be addressed to Douglas H. Wedell, Department of Psychology, University of South Carolina, Columbia, South Carolina 29208. Electronic mail may be sent via Internet to wedell@garnet.cla.sc.edu.

frequencies of 4, 3, 2, 1, and 1, respectively. Another group may be exposed to a negatively skewed distribution, with these same stimulus values presented with frequencies of 1, 1, 2, 3, and 4, respectively. The percentile rank function is negatively accelerated for the positively skewed distribution, with the percentile rank for Stimulus 3 being well above the 50th percentile. In contrast, the percentile rank function is positively accelerated for the negatively skewed distribution, with the percentile rank for Stimulus 3 being well below the 50th percentile. Insofar as individuals use the frequency principle in their judgments, mean rating functions will be similar in form to the cumulative frequency functions, producing strong contextual differences in mean ratings. Note that the differences in the mean ratings will be nonmonotonically related across contexts. For example, the difference in mean ratings for stimulus pair 1–3 will be greater than the difference in mean ratings for stimulus pair 3–5 in the positively skewed distribution. This is because the differences in percentile ranks is much greater for the former pair. The opposite relationship will then hold for negatively skewed distribution.

Krumhansl's (1978) distance–density formulation implies that this type of frequency (or density) effect occurs at the representational level and thus affects interstimulus similarity. However, an alternative formulation of the frequency principle as reflecting a tendency to assign equal numbers of stimuli to each rating category (Parducci, 1965) is also consistent with an interpretation of density effects occurring at an output or response stage. The fact that the magnitude of these effects has been found to depend on the number of rating categories (Parducci & Wedell, 1986; Wedell & Parducci, 1988; Wedell, Parducci, & Lane, 1990) further challenges the primacy of density effects.

Although Krumhansl (1978, 1982) provided intriguing post hoc applications of the distance–density principle to similarities data, the fundamental premise that changes in stimulus densities produce corresponding changes in judged similarity has yet to be experimentally confirmed. Indeed, two extensive experimental investigations in this area have generally not found density effects on measures related to pairwise dissimilarity (Corter, 1987; Mellers & Birnbaum, 1982). Corter (1987, 1988), in particular, has argued that without evidence that manipulations of stimulus densities affect pairwise similarity measures, the validity of applying the distance–density model to nonexperimental data is questionable at best. Similarly, DeSarbo, Manrai, and Burke (1990) have pointed out that the usefulness of psychometric scaling methods based on the distance–density model will depend on experimental evidence outlining the conditions under which density manipulations can be demonstrated to alter similarities.

The present article addresses this issue by presenting a constructive–associative model of unidimensional context effects that distinguishes different conditions under which manipulations of stimulus densities may alter the structure of unidimensional similarities. Assuming a monotonic function relating the similarity measures to the underlying similarity structure, I will interpret density manipulations that significantly alter the ordering of pairwise similarities as operating at the representational level. The experiments reported here explored how similarity relationships may depend on task constraints. Specifically, previous researchers in this area (Corter, 1987; Mellers & Birnbaum, 1982) have used an experimental paradigm in which pairwise ratings of psychophysical stimuli were made under conditions in which (a) there was no previous exposure to the stimulus distribution and (b) the two members of each pair were presented simultaneously. As Krumhansl (1988) has pointed out, such conditions may not be particularly representative of judgment situations typically encountered in which exemplars from a previously experienced set of stimuli must be retrieved from memory for comparison. The present experiments varied (a) whether or not participants rated the stimuli on a unidimensional scale prior to pairwise ratings and (b) whether or not members of a pair were separated by a delay. As described in the model presented below, different patterns of contextual dependence across these task conditions imply different processes underlying the construction of unidimensional similarity.

Constructive–Associative Model of Unidimensional Judgment

The issue of whether density manipulations affect similarity in large part depends on how density effects are conceived as operating on unidimensional judgments, where they are regularly observed. Figure 1 presents a model of unidimensional judgment that consists of both constructive and associative processes. The constructive process is modeled after range–frequency theory (Parducci, 1983) and may be conceived as the rapid generation of a judgmental category based on an implicit comparison of the value of the target stimulus with values of contextual stimuli. The associative process is modeled in terms of the retrieval of prior categorical associates and use of these as a basis for judgment.

The top row of boxes in Figure 1 illustrates the representations of the contextual distribution, the stimulus, and prior stimulus–response associates, respectively. The physical stimulus is assumed to give rise to a scale value S_i that represents its central tendency on the dimension of judgment. The constructive judgment process requires the retrieval of the distribution of contextual stimulus representations from previous trials, as shown in top left box. The range–frequency process consists of evaluating the location of S_i within the retrieved contextual distribution according to range and frequency principles and will be described in greater detail in the next section. The retrieval of associated categories is represented in the top right box as connections between a node representing stimulus scale value (S_i) and nodes representing the response categories assigned to that value (C'_1, C'_2, C'_3). Stimulus and response are connected through links to episodic contextual nodes (the small circles), which may operationally correspond to the different experimental trials. For example, in Figure 1, S_i has been associated with response category C'_2 on one trial and with response category C'_3 on three trials.

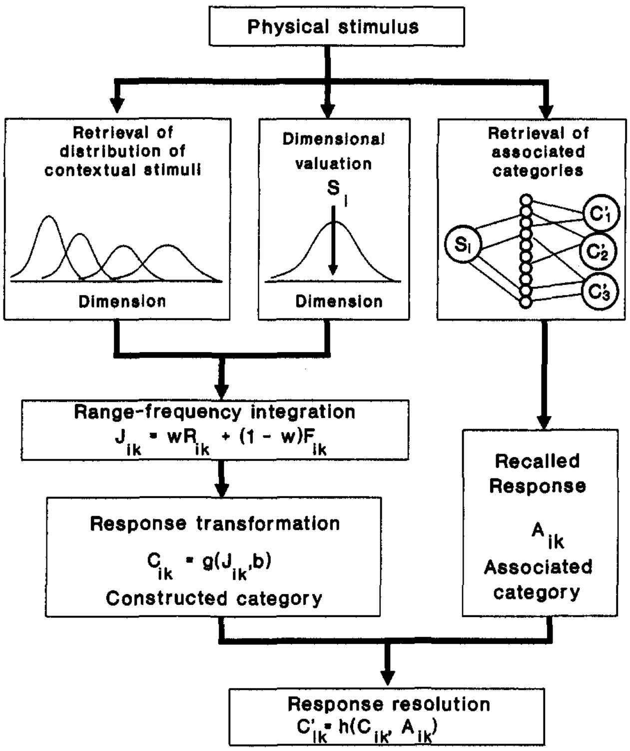

Figure 1. Constructive–associative model of unidimensional judgment. Symbols represent context independent scale values (S), range values (R), frequency values (F), internal judgments (J), range–frequency constructed categories (C), associated category (A), and selected category (C').

The model assumes that both the stimulus-driven constructive process and the response-driven associative process may be used to produce judgments on a trial. On early trials, when there have been few categorical responses, the constructive process will be primarily responsible for generating judgments. As more and more responses are made, a well-formed associative structure will be readily available so that the participant may simply retrieve the category most strongly associated with the value of the target stimulus and use this as a response. Thus, the constructive process may be bypassed altogether, or when a discrepancy arises between the outputs of the two processes, some type of response resolution may take place (pictured in the bottom box of Figure 1).

The judgment process depicted in Figure 1 applies to unidimensional judgments of successively presented stimuli. Judgments of the dissimilarity of pairs of unidimensional stimuli can then be conceived as a monotone function of the absolute differences in the dimensional values of those stimuli. Whether manipulations of contextual densities affect the ordering of pairwise dissimilarities will depend on which stimulus values are being compared (i.e., context-independent or context-dependent values). The fundamental premise of the constructive–associative theory is that different types of stimuli and different task constraints can lead to comparisons of the stimuli at different levels of representation. Context-independent dissimilarity judgments arise from direct comparison of the scale values (S_is). Context-dependent dissimilarity ratings result from comparison of either context-dependent categorical associates (A_{ik}s) or contextually constructed judgments (J_{ik}s). Before exploring the implications of the model for the particular tasks used in the present experiments, I will describe the constructive and associative processes in more detail.

Range–Frequency Theory

The effects of manipulating the distribution of stimuli along the dimension of judgment on category ratings are well described by Parducci's (1963, 1983) range–frequency theory of judgment. According to range–frequency theory, the judgment of a stimulus represents a compromise between a range and a frequency principle. For these principles to operate, it is assumed that presentation of the physical stimulus elicits a process whereby the value of the stimulus is located on the relevant dimension as well as a process whereby dimensional values of contextual stimuli are retrieved. Like Thurstone's (1927) conceptualization, each dimensional scale value (S_i) can be conceived as the mode of a distribution of dimensional values elicited over occasions. Furthermore, these scale values are generally assumed to be context independent, the contextual dependencies in judgments arising from the range–frequency processes that compare the value of the stimulus being judged to the values of contextual stimuli.

The range principle reflects a tendency to evaluate the stimulus in terms of the proportion of the contextual range lying below it:

$$R_{ik} = (S_i - S_{\min,k})/(S_{\max,k} - S_{\min,k}), \qquad (1)$$

where R_{ik} is the range value of stimulus i in context k, S_i is the context-independent scale value of the target stimulus on that dimension, and $S_{\max,k}$ and $S_{\min,k}$ correspond, respectively, to the scale values of the maximum and minimum stimuli that define the stimulus context. Because range values represent a linear transformation of context-independent scale values (S_i), differences in range values also will be a linear transformation of context-independent scale values. Thus, the ordering of pairwise dissimilarities based on a monotonic transformation of differences in range values will not vary across contextual distributions.

The frequency principle reflects the tendency to evaluate the stimulus in terms of the proportion of contextual stimuli lying below it:

$$F_{ik} = (r_{ik} - 1)/(N_k - 1), \qquad (2)$$

where F_{ik} is the frequency value of stimulus i in context k, r_{ik} is its rank in the set of contextual stimuli, and N_k and 1 correspond, respectively, to the ranks of the maximum and minimum stimuli that define the stimulus context. Because frequency values represent a nonlinear transformation of the context-independent scale values, the ordering of pairwise dissimilarities based on a monotonic transformation of differences in frequency values will likely vary across contextual sets. For example, consider stimulus *B*, whose scale value lies between the scale values for stimulus *A* and stimulus *C* on the dimension of judgment. By including contextual stimuli with values between *A* and *B*, the frequency values for *A* and *B* will differ by a greater amount than those for *B* and *C*. If instead, contextual stimuli are included with values between *B* and *C*, the reverse ordering of differences will occur. Thus, to the extent that similarity judgment is based on differences in frequency values, disordinal effects of density manipulations should be observed.

The internal judgment (J_{ik}) of stimulus *i* in context *k* reflects a compromise between range and frequency tendencies as represented by a weighted average of the corresponding values:

$$J_{ik} = wR_{ik} + (1 - w)F_{ik}, \qquad (3)$$

where *w* is to the weighting of the range value. In several experiments, *w* has been inferred to be close to 0.5 (Birnbaum, 1974; Parducci, 1965; Parducci & Perrett, 1971; Wedell et al., 1989). Finally, the overt category assigned to the stimulus is assumed to reflect an equal-interval partitioning of the internal judgment scale:

$$C_{ik} = g(J_{ik}, b), \qquad (4)$$

where C_{ik} is the category rating of stimulus *i* in context *k*, *b* is the number of rating categories, and g is a function that linearly partitions the judgment scale. Thus, for example, given an internal judgment of 0.15 and a five-category scale, the stimulus would be assigned to the first rating category because it falls into the interval 0.00 to 0.20. When the number of categories is five or more, mean ratings are well predicted by a linear response function (Parducci, 1983; Parducci & Wedell, 1986).

Retrieval of Associated Categories

As described above, range–frequency theory ignores any influence of the stimulus–response associates that presumably are formed during a judgment session. However, vast literature on category learning suggests that stimulus–response associates play an important role in the assignment of a stimulus to a particular response category (for reviews see Medin & Heit, in press; Medin & Smith, 1984). For example, exemplar-based models assume that categorization is determined by the associative strengths between the target and exemplars from the relevant set of categories (Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1992).

Within the literature on unidimensional judgments, the issue of stimulus–response associates has been most important in the study of transfer effects that occur when the distribution of contextual stimuli is changed within an experimental session. If the stimulus distribution on which range–frequency processes operate consists of only the most recent (e.g., last 15) trials (cf. Parducci & Wedell, 1986; Wedell, 1984), then the rating scale should shift rapidly and completely following a shift in the contextual distribution. However, this is not always the case. Transfer experiments have yielded mixed results: Sometimes shift of scale is rapid and nearly complete (Johnson, 1949, Tresselt, 1947), and sometimes it shifts only partially (Di Lollo & Casseday, 1965; Haubensak, 1992; Parducci, 1956). The situation is further complicated by the fact that individuals appear to differ in the degree to which they show transfer (Parducci, 1956; Wedell, 1984). These results suggest that in a transfer situation, a conflict may arise between the constructed category produced by a range–frequency process operating over recent postshift trials and the retrieved categorical associate that reflects judgments from earlier preshift trials.

For the present purposes, it is not necessary to specify the processes guiding retrieval. Instead, the model simply assumes that for each judgment trial, the scale value of the stimulus is associated with the overt categorical response assigned to that stimulus. As this associative network becomes better established over judgment trials, an associated category, A_{ik}, will be retrieved automatically at the time of judgment. If the contextual distribution remains constant across the course of the experiment, then the constructed and associated categories are likely to be the same. When a conflict does arise between the two, the model assumes that response resolution takes place, which can be represented as follows:

$$C'_{ik} = h(C_{ik}, A_{ik}), \qquad (5)$$

where C'_{ik} is the overt response to stimulus *i* in context *k*, and *h* is a function for selecting a category from the inclusive range C_{ik} to A_{ik}. It is then this overt response, C'_{ik}, that is encoded with the corresponding scale value for that trial.

The reason for raising the distinction between a constructed and associated category in the present setting is that contextual effects on pairwise similarity may result from comparisons of either of these two values, because the associated categories reflect the range–frequency effects. The experiments reported below were designed to distinguish between these two bases for the contextual dependence of similarity judgments, as described in the next section.

Implications for Similarity Judgments Across Task Situations

Three Bases for Unidimensional Similarity

The dependent measure of primary interest in the present set of investigations was a direct rating of pairwise dissimilarity. As described earlier, whether the ordering of pairwise dissimilarities differs across contexts will depend on whether context-independent or context-dependent values of the stimuli are compared. A failure to find disordinal density effects on pairwise dissimilarity judgments for unidimensional stimuli implies that unidimensional similarity is a function of context-free scale values, which can be denoted as follows:

$$D_{ij,k} = f_k(\mid S_i - S_j \mid), \qquad (6)$$

where $D_{ij,k}$ is the rated dissimilarity of stimulus pair ij in context k, and f_k is a positive monotone function on the difference in scale values of the two stimuli. Although the monotone function is allowed to vary across contexts, and hence affect dissimilarity judgments, this type of variation will not affect the ordering of pairwise dissimilarities. Note that because range values are a linear transformation of scale values, the context-independent dissimilarity model of Equation 6 could also be represented in terms of the absolute difference in range values R_{ik} and R_{jk}.

The constructive–associative model describes two different ways for disordinal context effects on pairwise dissimilarities to occur. The first is that pairwise dissimilarity judgments may be a monotonic function of the absolute difference in the constructed categorical judgments of the two stimuli:

$$D_{ij,k} = f_k(\mid J_{ik} - J_{jk} \mid). \tag{7}$$

Although Equation 7 uses the internal judgments, it could also use the constructed internal judgments (e.g., C_{ik}). If one assumes that the number of covert categories is relatively large (e.g., seven or more), the categorical values approximate a simple linear transformation of the internal judgments, and thus, these possibilities are formally equivalent. If (covert) categorization is based on only two or three categories, then similarity judgments based on the C_{ik}s will tend to differ from those based on J_{ik}s, although both will be context dependent. It is important to note that Equation 7 will only result in disordinal context effects on dissimilarities when $w < 1.0$, so that the frequency principle receives some weight. When $w = 1.0$, Equation 7 reduces to Equation 6 and no disordinal context effects will result.

A second way disordinal density effects could occur for pairwise dissimilarities would be to base pairwise judgments on the response categories associated with each stimulus:

$$D_{ij,k} = f_k(\mid A_{ik} - A_{jk} \mid). \tag{8}$$

This model of dissimilarity judgments differs from that of Equation 7 in that a set of stimulus–response associations must first be developed for context effects to occur. Thus, for example, when no stimulus–response associations have been built up, dissimilarity judgments may default to a context-independent process (Equation 6). Note also that disordinal effects on dissimilarities occur by using Equation 8 only when prior judgments are based on a range–frequency judgment rule that gives some weight to the frequency principle (i.e., $w < 1.0$).

Mapping the Model Onto Different Experimental Situations

On the basis of the constructive–associative model of judgment and the results of previous research, Figure 2 outlines the necessary and sufficient conditions for density effects that fail to occur in pairwise judgments across four different experimental situations. These experimental situations represent the factorial combination of (a) whether or not participants have made prior judgments of the contextual set of stimuli along the appropriate dimension (thus establishing context dependent stimulus–response associations), and (b) whether or not the members of the pairs are separated by a delay (thus requiring that they be held in memory).

Delay between pair members	Prior single stimulus judgment trials: No	Prior single stimulus judgment trials: Yes
No	No density effects IFF stimuli are commensurable and context effects have low priority	No density effects IFF stimuli are commensurable and context effects have low priority and associates ignored when stimuli present
Yes	No density effects IFF stimuli are commensurable and context effects have low priority and context-free values held in memory	No density effects IFF stimuli are commensurable and context effects have low priority and context-free values held in memory and associates ignored when stimuli in memory

Figure 2. Theoretical conditions leading to a lack of density effects on pairwise similarity as defined by the model under the four experimental situations generated by factorially manipulating prior single-stimulus judgment trials and delay between pair members. IFF denotes if and only if.

The top left cell of Figure 2 represents the situation that has been most extensively explored in previous experimental research. In this situation, the conjunction of two conditions is necessary for context effects not to occur. First, the stimuli must be commensurable or directly comparable. Incommensurable stimuli can be defined as those that have representations so different that they cannot be directly compared. If stimuli are incommensurable, then each must be judged implicitly or explicitly on the relevant attribute dimensions, with these judgments forming the basis of the comparison. Because these judgments are context dependent, the dissimilarity ratings on the basis of these judgments will also be context dependent.

The role of stimulus commensurability is nicely illustrated by experiments conducted by Mellers and Birnbaum (1982). In their Experiment 2, individuals rated the perceived difference in the darkness of pairs of dot patterns presented simultaneously on a page. Frequency manipulations did not produce disordinal effects on difference ratings, implying a low priority of contextual processing. However, in their Experiment 3, individuals made cross-modality difference judgments, rating whether a dot pattern was darker than a circle was large. Under these conditions, disordinal context effects on difference ratings emerged. The ordering of pairwise differences varied across contexts in a manner consistent with the range–frequency effects on the single-stimulus ratings for each dimension. These effects may be interpreted as resulting from the incommensurability of the stimuli. Because dot patterns could not be directly compared with circles on a common dimension, implicit context-dependent judgments of each stimulus may have served as the bases for difference ratings (e.g., as in Equation 7).

The failure to find significant density effects in experiments corresponding to the top left cell of Figure 2 (Corter, 1987, Experiments 1, 2, and 3; Mellers & Birnbaum, 1982, Experiment 2; Roberts & Wedell, 1994) implies that the stimuli in those studies were commensurable and the context effects had low priority. Although the incommensurability issue is an important one, the experiments reported in the present article used only commensurable stimuli. Commensurability was ensured by having participants make pairwise dissimilarity ratings of unidimensional psychophysical stimuli (e.g., squares that differed in size only).

The top right cell of Figure 2 corresponds to the situation in which both members of the pair are simultaneously present at the time of judgment, but they have been rated in a prior unidimensional judgment task. These prior ratings would be expected to show the usual density effects, and thus in this situation a set of context-dependent stimulus–response associations is available. If dissimilarity ratings are based on a comparison of these associates (Equation 8), then the ordering of pairwise dissimilarities would be expected to vary across contexts. Thus, a failure to find density effects for experiments described by the top right cell of Figure 2 would imply the two conditions of the top left cell, as well as a third condition, that the context-dependent categorical associates are ignored when the stimuli are simultaneously present.

The bottom left cell of Figure 2 corresponds to the situation in which members of the pairs are separated by a delay but have not been previously rated on a unidimensional scale. Because no categorical associates have been formed in this situation, any observed context effects would reflect a comparison process operating on the constructed categories (Equation 7). A lack of density effects for experiments described by this cell would imply the two conditions of the top left cell, as well as a third condition, that context-free scale values are held in memory during the delay. If this last condition cannot be met, then it is assumed that context-dependent values are covertly constructed and held in memory for comparison in the similarities task.

Finally, the bottom right cell of Figure 2 corresponds to the situation in which members of the pairs are separated by a delay and also have been rated previously on a unidimensional scale. Density effects in this cell could arise from either constructive or associative processes. A lack of density effects for experiments falling into this cell would imply that the three conditions of the bottom left cell of Figure 2 hold as well as a fourth condition, that categorical associates are ignored even when stimuli must be held in memory.

Meaningful Patterns of Results

Table 1 presents the six interpretable patterns of results that follow from the analysis presented in Figure 2. First, if for a set of stimuli, no density effects occur in any of the four cells, then all of the conditions implied in those cells are assumed to hold. (Naturally, confidence in the implications derived from retaining the null hypothesis would depend on the degree of power built into the experimental design.) This pattern would provide the strongest support for the response-bias hypothesis, suggesting that contextual effects in unidimensional judgment (for that domain of stimulus) are strictly a response artifact. A response bias interpretation of context effects implies that no aspect of the stimulus representation is affected by manipulation of context, but rather response tendencies, such as equalization of category use, produce these effects. Because contextual processes are strictly tied to response generation in the unidimensional rating task, they should not form the basis for dissimilarity ratings and hence no disordinal context effects on dissimilarities should be observed.

Second, if a set of stimuli shows density effects only when there has been both a delay between pair members and prior single-stimulus judgment trials, then all of the implied conditions hold except the last—that associates are ignored when stimuli must be held in memory. This pattern represents the second strongest version of the response-bias interpretation of context effects in unidimensional judgment

Table 1
Six Interpretable 2 × 2 Patterns of Effects Corresponding to Experimental Situations Defined in Figure 2

Pattern	Interpretation
00 00	1. Context-independent default
00 0+	2. Associative-based default only under memory constraints
0+ 0+	3. Associative-based default
00 ++	4. Constructive-based default under memory constraints
0+ ++	5. Associative-based default with constructive-based default under memory constraints
++ ++	6. Constructive-based default

Note. 0 indicates no density effects on pairwise similarities; + indicates density effects on pairwise similarities.

in that similarity is affected only through response associates and only when memory for the stimuli is constrained.

Third, and again consistent with an associative basis for similarity (Equation 8), density effects may occur only when prior response associates have been formed (the right cells of Figure 2). This pattern implies that the conditions described in the left cells of Figure 2 are all true, but that response associates, once formed, cannot be ignored in making judgments of similarity.

Fourth, and in line with a constructive basis for context effects on similarity judgments (Equation 7), density effects may occur only when there is a delay between members of the pairs. This pattern implies that context effects on similarity are overridden when simultaneous presentation makes direct comparison of scale values possible; however, context-free values are not easily held in memory and, therefore, judgments are based on the constructed categories in the delay conditions. The linking of density effects to situations in which stimuli must be held in memory is consistent with the finding that density effects have been found for confusion matrices generated from identification tasks (Appleman & Mayzner, 1982; Corter, 1987; Krumhansl, 1978) and sorting tasks (Roberts & Wedell, 1994). In such tasks, the presented stimulus must be compared with stimulus representations residing in memory. Thus, if memory encoding or retrieval is context dependent, then density effects may emerge that otherwise might be absent when there are no memory constraints involved.

Fifth, density effects may be observed either when there is a delay between pairs or when a prior response scale has been formed. Thus, even though contextual processing would be assumed to have low priority in unidimensional judgment, density effects generally would be observed whenever context-dependent response associates were available or whenever stimuli had to be held in memory for comparison. This pattern might help to reconcile Krumhansl's (1978) assumption of pervasive density effects on similarity with the lack of density effects on pairwise judgments of visual stimuli under simultaneous presentation conditions (Corter, 1987; Mellers & Birnbaum, 1982). It may well be that simultaneous presentation of previously unjudged stimuli is a rather atypical situation. Moreover, simultaneous presentation may not be a viable option for many types of stimuli, such as auditory or verbal stimuli.

Sixth, for some stimulus domains, density effects may be observed under all four conditions. This would imply that either stimuli are incommensurable or context effects have high priority (or both). Although Mellers and Birnbaum (1982, Experiment 3) explored pairwise judgments for incommensurable stimuli only in conditions corresponding to the top left cell of Figure 2, the model implies that these same density effects would necessarily occur in the other three cells. Although the relatively few experiments conducted in the no-delay, no-prior-judgment cell did not yield density effects on pairwise judgments of commensurable stimuli, this does not mean that contextual effects have low priority in all stimulus domains. It is entirely possible that the priority of contextual effects is tied to the stimulus domain.

Finally, it should be noted that the framework presented in the present article is testable in the sense that of the 16 possible patterns of results for the design of Figure 2, 10 of these would be incompatible with the model. For example, if density effects occurred for a stimulus domain in the no-delay, no-prior-judgment cell, then they must also occur in the other three cells. In more general terms, a pattern is incompatible with the model if no density effects occur in a cell that contains conditions found in a cell for which density effects have occurred.

Overview of Experiments

The experiments described in the present article followed the design of Figure 2 and attempted to determine which of the six interpretable patterns of Table 1 apply to psychophysical judgments for specific stimulus domains. The dependent measure of primary interest was pairwise ratings of dissimilarity. The stimuli were squares that varied in size in Experiments 1 and 2 and dot patterns that varied in number of dots in Experiment 3. In Experiment 1, pairwise dissimilarity ratings of square size were made either prior to or after single-stimulus judgments of size. This manipulation tests for the effects of building contextually dependent categorical associates. The two squares making up a pair were either presented simultaneously on the screen or separated by a delay to test whether the delay would lead to construction of context-dependent values used for the dissimilarity judgments. Experiment 2 used squares as stimuli as well, but varied the spacing of stimuli. Because results for square judgment appeared to contradict some earlier results for judgments of dot patterns, Experiment 3 replicated the basic design of Experiment 1 using dot patterns as stimuli. Experiment 3 was also designed to test specific predictions from Krumhansl's (1978) distance–density theory concerning density effects on judgments of similarity of a stimulus to itself and on asymmetry of similarity judgments.

Strategy for Model Fitting

A major purpose of the present article is to determine the degree to which the range–frequency model developed for describing single-stimulus category ratings can be extended to explain pairwise dissimilarity ratings. Therefore, differences in pairwise dissimilarity ratings across contexts will be modeled strictly in terms of changes in the range weighting parameter, w.

The range–frequency model of Equations 1–3 describes how properties of the distribution operate on context-independent scale values (S_is) to create internal judgments (J_{ik}s). The model does not specify the psychophysical function that determines how physical values are translated to context-independent scale values. For the stimulus materials used in the present set of experiments, this function can be represented as

$$S_i = m(\phi_i), \tag{9}$$

where Φ_i is a measure of the physical attribute of stimulus i and m is a monotone function. In the ideal situation, the psychophysical function would be invariant across manipulations of presentation mode and judgment task. In the present set of experiments, the psychophysical function was found to depend on presentation mode and judgment task for squares but not for dot patterns. In modeling data in each experiment, however, the psychophysical function was always constrained to be the same across contexts.

In addition to the psychophysical function, response functions are required to translate internal judgments to overt responses. For single-stimulus ratings, the range–frequency model typically assumes that mean ratings are a linear function of internal judgments. Parducci (1983) has further recommended setting the additive constant to the lowest numeral on the rating scale and the multiplicative constant to the range of numerals on the rating scale, so that one can unambiguously interpret the range–frequency contextual parameters. This procedure was followed in fitting the rating data for Experiments 1–3. Because the single-stimulus category ratings were made using a 9-point scale, the response function was given by

$$\bar{C}_{ik} = 1 + 8J_{ik}. \tag{10}$$

The response function for the dissimilarity ratings depends on how distance is related to dissimilarity. Nosofsky (1992) has argued that the nonlinear relationship between distance and similarity is well described by either an exponential decay or a Gaussian similarity function. Shepard (1987) has further proposed that exponential decay functions represent a universal law of stimulus generalization. Given the success of these models, an exponential decay similarity function was used as the response function in the present set of experiments. Mean dissimilarity ratings were modeled as a linear function of the inverse exponential distance calculated from the internal judgment scale:

$$\bar{D}_{ij,k} = a - b[\exp(-c \mid J_{ik} - J_{jk} \mid)], \tag{11}$$

where a and b are the constants defining the linear transformation and c is a sensitivity parameter. When $w = 1$, Equation 11 corresponds to Equation 6, with dissimilarity judgments based on differences in context independent values. When $w < 1$, disordinal contextual effects on mean dissimilarity ratings will generally result. Because the stimulus distribution for any prior rating task was always the same as that used in the dissimilarity rating task, Equation 11 can be used to model either constructive or associative bases of context effects (i.e., Equations 7 or 8).

In each experiment, the predicted mean ratings generated from Equation 10 for the single-stimulus task and Equation 11 for the pairwise dissimilarity task will be compared with empirical mean ratings to examine how well the range–frequency model accounts for contextual effects in the two tasks. In addition, the estimated internal judgments (J_{ik}s) from Equation 11 will be compared with scale values derived from nonmetric multidimensional scaling (MDS). The MDS model provides inferred scale values based on order information alone, without specific assumptions about the psychophysical function or response function. A close correspondence between the theoretically derived judgment scale (J_{ik}s) and the theoretically derived MDS values will provide further support for the range–frequency model. Finally, Experiment 3 also includes tests on density effects predicted by Krumhansl's (1978) distance–density model. Predictions from the distance–density and range–frequency models will be compared.

Experiment 1: Effects of Prior Ratings and Intrapair Delay on Dissimilarity Judgments of Squares

Judgments of sizes of squares have been extensively studied within range–frequency theory (Parducci, 1963, 1965; Parducci, Knoble & Thomas, 1976; Parducci & Perrett, 1971; Parducci & Wedell, 1986). These studies converge on following basic findings: (a) The context-free scale values (S_i) inferred from range–frequency fits to data tend to be a linear function of square width, (b) the discriminability scale inferred from Thurstonian techniques tends to be a logarithmic function of square width, (c) when stimulus spacings are manipulated, the inferred weighting of frequency values $(1 - w)$ tends to range from about 0.4 to 0.5, regardless of number of categories, and (d) when stimulus frequencies are manipulated, the inferred value of $(1 - w)$ varies inversely (from as much as 0.80 to as little as 0.10) as number of categories increases.

With regard to findings c and d, the density manipulations in Experiment 1 consisted of unequal spacings of contextual stimuli along the dimension of judgment, rather than unequal frequencies to create large and stable density effects (regardless of the number of categories participants might use in covert categorization). With regard to findings a and b, it was unclear whether similarity judgments would follow the linear scale, logarithmic scale, or some intermediate scale. Most of the conditions reported here were based on linear spacing; however, the results prompted some limited experimentation with stimuli assuming an intermediate scale.

Figure 3 graphically depicts the form of the expected density effects on unidimensional judgments and how those effects could be manifest in pairwise dissimilarity judgments. The stimulus values making up positively and negatively skewed distributions in Figure 3 were the same as those actually used in Experiment 1. The three panels represent the three data analytic techniques used for understanding potential contextual effects in each experimental condition.

The left panel presents range–frequency predictions of magnitude ratings on a 9-point scale for squares drawn from a set of 25 sizes (with each successive size representing a constant increment in physical width). The predictions use Equations 1–3 and 10, making the following assumptions: (a) The end stimuli define the minimum and maximum values, (b) range and frequency principles are equally weighted (i.e., $w = .5$), and (c) context-independent scale values vary linearly with the physical widths of the squares.

Predicted rating functions are shown for stimuli making

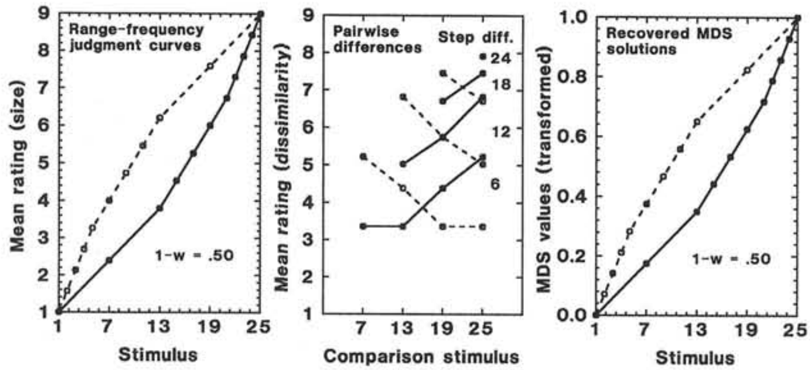

Figure 3. Theoretical predictions for context-dependent density effects on dimensional and pairwise ratings generated by the range–frequency model. Left panel represents theoretical functions for single-stimulus rating; middle panel represents theoretical functions for pairwise ratings; right panel represents theoretical functions for inferred multidimensional scaling values. MDS = multidimensional scaling.

up positively and negatively skewed distributions. The two distributions share five common square sizes, including the end stimuli, so that by Assumption a, the range functions are the same for the two distributions. The density manipulation consists of interspersing contextual stimuli between Squares 1 and 13 to create positive skewing and between Squares 13 and 25 to create negative skewing. Thus, manipulation of frequency values results in a negatively accelerated rating function for the positively skewed set and a positively accelerated function for the negatively skewed set.

As described in Equations 6–8, judgments of pairwise dissimilarity are assumed to be monotonically related to the perceived differences in the values of the two stimuli. The middle panel of Figure 3 presents the pattern that would be expected if pairwise dissimilarities were based on the exponential decay similarity function of Equation 11, with $a = 9$, $b = 8$, and $c = 2$. The rating functions correspond to predicted dissimilarity ratings for target-stimulus pairs that are separated by either 6, 12, 18, or 24 physical steps. The slopes of the rating functions for positive and negative sets are of opposite sign, indicating differences in the ordering of pairwise judgments. For example, the 1–7 pair is rated as more dissimilar than the 19–25 pair for the positively skewed distribution, but this ordering is reversed for the negatively skewed distribution. If pairwise dissimilarity ratings were based on context-independent scale values, the functions for positive and negative sets would be parallel.

Finally, the right panel of Figure 3 presents the scale values for the five target stimuli resulting from nonmetric MDS of the 55 pairwise difference ratings of stimuli for each contextual set, with scale values linearly transformed to range from 0 to 1 for each solution. As can be seen, the MDS technique recovered the form of the unidimensional rating scales, and when the range–frequency model was fit to these scale values, the same value of $1 - w$ was recovered.

Method

Participants. Participants included 174 University of South Carolina undergraduates participating in partial fulfillment of psychology course requirements, who were randomly assigned to one of eight between-subject conditions. Between 21 and 24 participants took part in each condition.

Following the design of Figure 2, the general experimental design consisted of a 2 × 2 × 2 factorial combination of distribution (positive or negative skewing), task order (single-stimulus ratings first or pairwise ratings first), and delay (0- or 3-s delay between-pair members in the pairwise task), all manipulated between participants. In the single-stimulus task, the dependent variable was the mean ratings (on a 9-point scale) of the sizes of five target stimuli common to the two distributions. In the pairwise task, the dependent variable was the mean ratings (on a 9-point scale) of the pairwise dissimilarities among the five target stimuli (i.e., altogether 10 different pairs).

Apparatus. Microprocessors (IBM Model PS2 50Zs with video graphics array [VGA] color displays) were used to present instructions, display stimuli, and collect responses. Stimuli were displayed in a VGA graphics mode (640 × 480 square pixels) as yellow squares on a blue background.

Stimuli. The main sets of stimuli were drawn from a set of 25 squares, varying by 10-pixel increments from 5 to 245 pixels in width (1 pixel = .3125 mm). The positively skewed set consisted of sizes 1, 2, 3, 4, 5, 7, 9, 11, 13, 19, and 25; the negatively skewed set consisted of sizes 1, 7, 13, 15, 17, 19, 21, 22, 23, 24, and 25.

For the single-stimulus rating task, each of the 11 squares in a set was presented once in a random order during preview trials, and then five times more in a block randomized order during the experimental trials for a total of 66 trials. For the pairwise rating task, each square was paired with every other square (but itself) for a total of 55 different pairings. The 11 preview trials consisted of a random sample from the 55 pairs, with the restriction that the pair 1–25 (representing the largest difference) was presented. Each pair was presented in a random order; once in the first 55 experimental trials and once in the second 55 experimental trials. The member of each pair that appeared on the left of the screen was randomized for the preview and the first 55 trials. For the second set of 55 trials, the member of each pair that had previously appeared on the left was presented on the right side of the screen.

Procedure. Up to 5 individuals participated at a time, each seated at a separate terminal, spaced approximately 1 m apart. General instructions stated that participants would be asked to make judgments of squares that varied in size. Depending on the experimental conditions, instructions were then presented for either the single-stimulus or pairwise judgment task.

For single-stimulus judgments, participants were instructed to rate the sizes of squares on a 9-point scale, with verbal labels, 1 (*very very small*) and 9 (*very very large*). Participants entered a number for each square, corresponding to how large or small that square seemed to them in comparison with all the other squares. To encourage participants to use the full range of categories, instructions stated that the smallest square should be rated 1 and the largest square should be rated 9. The 11 preview trials were then presented at a self-paced rate. Each trial consisted of a square being presented in the middle of the screen, with the rating scale presented graphically at the bottom of the screen. Participants entered their ratings by pressing the appropriate keys (1–9). After a response in the appropriate numerical range was entered, the square was immediately cleared from the screen, but the response remained on the screen for an additional 0.5 s. The next trial began 1 s after entry of the previous response. After the preview trials, the participants were reminded of the task instructions and then were presented with the 55 experimental trials.

In the pairwise task, participants were instructed to rate how similar or dissimilar in size each pair of squares seemed to them on a 9-point scale ranging from 1 (*very very similar*) to 9 (*very very dissimilar*). To encourage participants to use the full range of categories, instructions stated that the most similar pairs should be assigned a rating of 1 and the most dissimilar pairs should be assigned a rating of 9. If participants had first participated in the single-stimulus task, they were also informed that the squares would be the same as those presented in the first task. Participants in the delay condition were told how the squares would be presented on the screen and were encouraged to pay close attention during the task to avoid missing the presentation of one of the squares. All of the participants were told that the preview set included the largest and smallest squares in the set. The 11 preview trials were then presented at a self-paced rate. Each trial consisted of squares being presented on the left and right sides of the screen. In the simultaneous presentation condition, the squares were presented at the same time and remained on the screen until the rating was entered. In the delay condition, a square was presented on the left side of the screen for 1 s, the screen was cleared for 3 s, a square was then presented on the right side of the screen for 1 s, and then the screen was cleared again. The screen remained blank (except for the response scale) until a response in the correct range was entered. This response appeared on the screen for 0.5 s. A new trial began 1 s after the previous response had been entered. After the preview, participants were reminded of the task instructions and then 110 experimental trials were presented.

Results

Ratings of size. Figure 4 presents the 9-point ratings of size when the squares were rated prior to dissimilarity judgments and after dissimilarity judgments (combining 0- and 3-s delay conditions). The effects of context were as predicted, with mean rating functions for the positively skewed sets displaced higher than and negatively accelerated relative to the functions for the negatively skewed sets. However, an unanticipated finding was that the rating functions for conditions following pairwise dissimilarity ratings were negatively accelerated relative to the corresponding functions for conditions in which single-stimulus size ratings occurred first. Because the distributional manipulation is the same, these differences may be attributable to a change of scale values or of the response function.

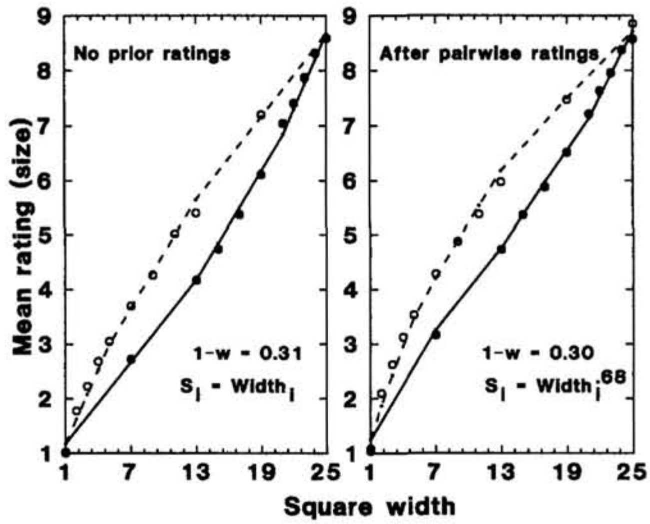

Figure 4. Range–frequency model fits generated from Equation 12 to 9-point ratings of size in Experiment 1. Introducing prior pairwise ratings results in a change in the psychophysical function but no change in the magnitude of context effects.

The range–frequency model was fit to the mean category rating data under the assumption that the psychophysical function relating physical values to scale values was a power function. Allowing S_i to be defined as a power function of physical width ($\Phi^p{}_i$) and using the response function of Equation 10 yielded the following equation:

$$\bar{C}_{ik} = 1 + 8\left(w\,\frac{\Phi^p{}_i - \Phi^p{}_{\min}}{\Phi^p{}_{\max} - \Phi^p{}_{\min}} + (1 - w)F_{ik}\right), \quad (12)$$

where $\Phi^p{}_{\min}$ and $\Phi^p{}_{\max}$ are constrained to be constant across contexts and p is held constant within prior task conditions. Equation 12 was fit to the data using least squares iterative nonlinear regression. The dashed and solid lines of Figure 4 represent the values predicted from the range–frequency model. When square size was rated first, only three parameters ($\Phi^p{}_{\min}$, $\Phi^p{}_{\max}$, and w) were needed to fit the data, with p fixed a priori at 1. When p was allowed to vary its estimated value was 1.01. The excellent fit of the range–frequency model validates prior results indicating the range values were a linear function of square width.

When square size was rated after dissimilarity judgments, the three-parameter model provided a poorer fit to the data. Allowing p to vary resulted in an excellent fit, with the fitted value of the power exponent equal to 0.68. For both sets of ratings, the inferred frequency weighting was about 0.30, which was somewhat smaller than expected. However, the differential curvature of the rating functions was still substantial enough to produce large disordinal effects if pairwise similarities were based on these ratings.

Analyses of variance (ANOVAs) were conducted to verify statistically the conclusions described above. First, a $2 \times 2 \times 2 \times 5$ ANOVA was conducted on participants' mean ratings of size for the data corresponding to the two panels of Figure 4, with distribution, delay, and order of tasks as the between-subject variables (each manipulated at 2 levels), and target stimulus as the within-subjects variable. A lack of a significant main effect of (or any interactions with) the delay variable ($p > .05$) provided justification for combining these data in Figure 4. The strong effects of the density manipulation were reflected in both the main effect of distribution, $F(1, 166) = 90.2$, $p < .001$, and the Distribution $\times$ Target Stimulus interaction, $F(4, 664) = 58.9$, $p < .001$. In accordance with range–frequency theory, only the quadratic component of the two-way interaction was statistically significant, $F(1, 166) = 132.7$, $p < .001$.

The difference in the inferred underlying scales of the middle and left panels of Figure 4 was reflected in a significant Task Order $\times$ Target Stimulus interaction, $F(4, 664) = 9.2$, $p < .001$. Trend analysis on the mean ratings of the five target stimuli for each task order revealed that although the linear component accounted for more than 98% of the systematic variance in each case, the two conditions differed in the quadratic trend. When single-stimulus ratings occurred first, there was no significant quadratic trend, $F(1, 82) = .8$, $p > .25$, but when single-stimulus ratings followed pairwise ratings, a significant quadratic trend emerged, $F(1, 84) = 57.4$, $p < .001$.

According to the range–frequency model, the emergence of a quadratic trend in the mean ratings for judgments made after pairwise ratings cannot be explained by range or frequency processes, but instead reflects either a change in the underlying psychophysical scale or in the response transformation function toward a more nonlinear, negatively accelerated function. Figure 4 demonstrates that a range–frequency model in which scale values become a negatively accelerated function of square width after dissimilarity rating provides a good description of the data. The negatively accelerated psychophysical function is consistent with the general Fechnarian principle underlying the discriminability scale for squares. As will be demonstrated in the following sections, it is also generally consistent with the inferred scales underlying pairwise dissimilarity judgments. Although Parducci (1982) has argued that the concept of a psychophysical law is obsolete, range–frequency theory has nevertheless been used to derive a context-independent scale that may be viewed as a psychophysical function (Birnbaum, 1974). The present results suggest that the psychophysical scale can change dramatically as a function of prior task. Hence, at least for size judgments, the form of the psychophysical function appears more labile than previously assumed.

Ratings of pairwise dissimilarities. As theoretically portrayed in the middle panel of Figure 3, the hypothesized density effects on dissimilarity imply that the same step difference between stimuli will be judged more dissimilar when the region between the stimuli is dense rather than sparse. Accordingly, mean ratings of dissimilarity should produce crossover interactions for same-step comparisons, so that the slopes of the functions are more negative for positively skewed than for negatively skewed conditions. In general, the mean ratings of dissimilarity for the 10 pairs of target stimuli portrayed in each of the panels of Figure 5 exhibit the predicted interactions. For example, target pair 1–7 was rated more dissimilar when presented in the positively skewed context in which values of four contextual stimuli were between 1 and 7 than in the negatively skewed context in which no contextual values were between the two stimuli. Conversely, target pair 19–25 was rated more dissimilar in the negatively skewed context than in the positively skewed context.

Separate two-way Distribution $\times$ Pair ANOVAs were run on the mean ratings of dissimilarity corresponding to the 6-, 12-, and 18-step comparisons of each panel. Figure 5 presents the significance levels for the interaction terms of these analyses. Because the number of participants and most aspects of the experimental method were approximately the same across panels, these significance levels can be used as rough, inverse indices of the effect size of the distribution manipulation when comparing the same-step differences across panels. Because the smaller step comparisons had more pairs and hence more degrees of freedom, significance levels should not be used to gauge relative effect sizes when comparing across different number of steps.

Examination of the pattern of significance leads to a number of conclusions. First, significant crossover interactions were obtained in all four conditions, reflecting the context-dependent nature of the dissimilarity judgments. Second, the contextual effects were most pervasive in the 3-s delay conditions, suggesting that requiring the participant to hold one member of the pair in memory led to a contextually dependent encoding of the stimulus representation. The magnitude of these effects were similar regardless of whether prior categorical associates were established. Third, the effects of context seem somewhat diminished in the 0-s delay conditions, especially when there were no prior ratings. These results may be interpreted within the constructive–associative model as reflecting a reduced tendency to construct context-dependent values when the stimuli are simultaneously present; however, this may be offset somewhat when context-dependent categorical associates are available at the time of judgment.

Separate two-way Distribution $\times$ Target Pair repeated-measures ANOVAs were conducted on the mean ratings of the 10 pairs for each of the four Delay $\times$ Prior Rating conditions shown in Figure 5. All four analyses revealed significant two-way interactions ($p < .05$), indicating that skewing affected similarity relations in each condition. A four-way, Distribution $\times$ Task $\times$ Delay $\times$ Target Pair repeated-measures ANOVA conducted on the mean ratings of the 10 target pairs revealed a significant two-way interaction between distribution and target pair, $F(9, 1494) = 17.3$, $p < .001$, $MSE = 1.187$, but none of the other interactions involving distribution and target pair were significant. As a more sensitive test, a measure of the average slope of the functions in each panel of Figure 5 was computed and subjected to a three-way Distribution $\times$ Task $\times$ Delay ANOVA. The effects of distribution interacted sig-

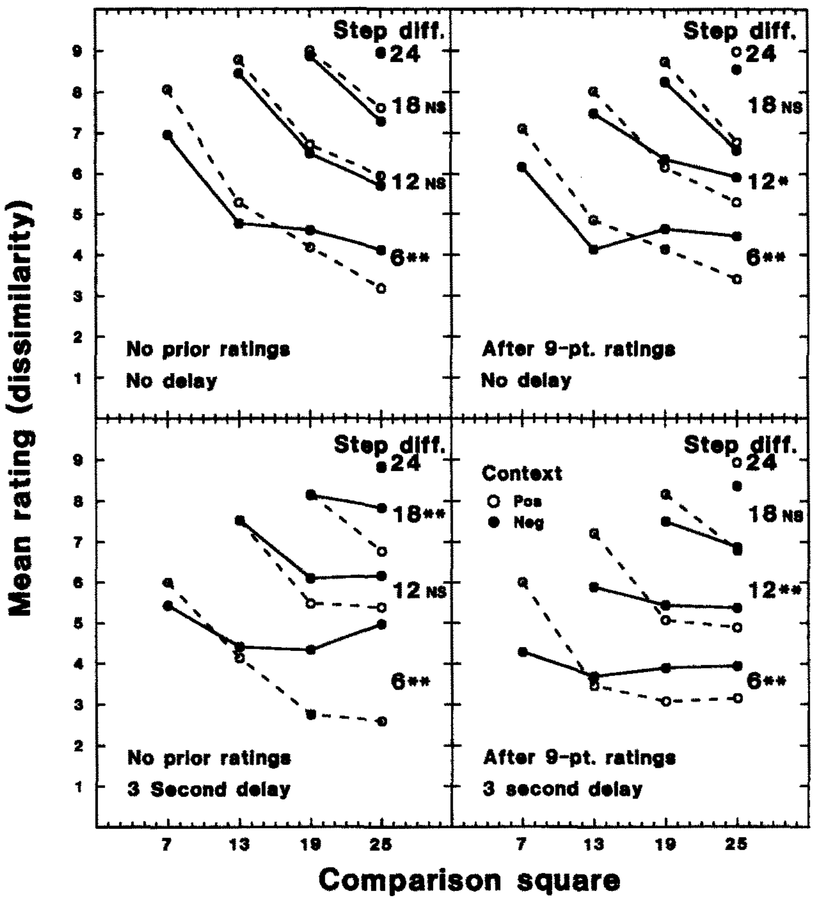

Figure 5. Mean ratings of pairwise dissimilarity for the 10 comparisons among the five target stimuli of Experiment 1. For each pair of functions in each panel, the significance level of the Distribution × Pair interaction term is reported at one of four levels: NS = nonsignificant ($p > .10$).

nificantly with delay, $F(1, 166) = 5.6$, $p < .05$, $MSE = .012$, reflecting the increased differences in slopes as a function of distribution in the 3-s delay conditions. The main effect of delay was also significant, $F(1, 166) = 29.101$, $p < .001$, indicating the generally steeper slopes of the functions in the 0-s delay conditions.

Unlike the hypothetical functions in the middle panel of Figure 3, the empirical functions of Figure 5 did not differ in the sign of their slopes, but rather all exhibited a general negative slope. This negative slope indicates that although equal differences in physical widths corresponded to roughly equal intervals along the subjective scale for ratings of size, this was not the case for scales underlying dissimilarity ratings. Instead, the inferred scales underlying comparisons of pairwise dissimilarity are more negatively accelerated, with the same increment in physical width corresponding to a greater increment in perceived dissimilarity when the squares being compared were small rather than large in size. This is evidenced by the fact that the mean ranked dissimilarity for the 1–7 pair was close to 5.5, whereas the mean ranked dissimilarity for the 19–25 pair (which corresponds to the same increment in physical width) was closer to 2. Furthermore, the steeper slopes of the functions in the 0-s delay versus the 3-s delay conditions indicate that this tendency was greater when squares were simultaneously present on the screen. These conclusions were supported by the results of Delay × Distribution × Target Pair ANOVAs run for the functions of top and bottom panels of Figure 5. In each analysis, the main effect of target pair and the Target Pair × Delay interaction were highly significant ($p < .001$).

Range–frequency modeling of dissimilarity data. To assess how well the range–frequency model can account for the dissimilarities data, the model was fit to dissimilarity ratings using a nonlinear iterative regression technique described below. The model's fit was assessed at two levels. First, it was evaluated by determining whether the range–frequency inferred scale values corresponded to scale values derived from nonmetric MDS analyses applied to the mean dissimilarity ratings of each contextual condition. The nonmetric MDS technique uses only ordinal information from the data to derive scale values and is not theoretically driven. A second way to evaluate the model is to determine how well the model predicted the dissimilarity ratings themselves.

A nonmetric MDS routine (Wilkinson, 1988) was applied to each triangular matrix of mean dissimilarity ratings consisting of 55 pairs of stimuli in each condition. Nonmetric scaling uses only the ordinal information from the judgment data and hence provides a method for understanding the contextually induced differences in the underlying similarity structure implied by the differences in the ordering of pairwise dissimilarities. In each analysis, a one-dimensional solution was selected on the basis of a priori considerations as well as the low values of stress for these solutions.[1] To compare positive and negative sets, scale values were linearly transformed to a 0–1 scale by subtracting the lowest value and dividing by the range of values. The MDS scale values for each contextual condition are shown in Figure 6 as the solid and open circles. Although there were relatively few reversals of the ordering of pairwise dissimilarities, there were enough to produce systematic context effects on scale values. These context effects on MDS derived functions were of the form predicted by range–frequency theory, with greater negative acceleration for the positively skewed distributions.

The range–frequency model was fit to the mean dissimilarity ratings shown in Figure 5 by using iterative nonlinear regression to estimate the parameters of the following equation:

$$\bar{D}_{ij,k} = a - b\exp[-c(w \mid \Phi^p_i - \Phi^p_j \mid /(25^p - 1) + (1 - w) \mid F_{ik} - F_{jk} \mid)]. \quad (13)$$

The five free parameters of Equation 13 were fit to the 20 data points within each panel of Figure 5 (i.e., each Prior Task × Presentation Mode condition). The only parameter allowed to vary across contexts was $1 - w$, the frequency weighting. Thus, all predicted differences between contexts were generated from the range–frequency model. The other four fitted parameters were the power exponent p describing the psychophysical function and the response parameters a, b, and c, corresponding to the exponential decay dissimilarity function of Equation 11. The relevant range was assumed to be the largest and the smallest squares, 25 and 1.

Figure 6 demonstrates a close correspondence between the scale values inferred from the range–frequency model of Equation 13 (dashed and solid lines) and the scale values inferred from the atheoretical MDS technique. In each panel, the range–frequency functions are generated from only two parameters, $1 - w$ and p. These are instructive in analyzing the data. First, note that contextual effects are small when there is no intrapair delay and no prior ratings, $1 - w = 0.08$. Providing prior ratings increases the contextual dependence, $1 - w = 0.19$. Furthermore, allowing a 3-s intrapair delay increases contextual dependence as well, with the weighting values of 0.26 and 0.25. These results indicate effects of both associative and constructive processes.

Second, the general negative acceleration of the functions is captured by the power parameter, p. Values of p indicate that prior rating did not affect the acceleration of the scale values but that delay did. The extremely low values of p in the no-delay condition begin to approach the logarithmic form of the discriminability function. The negative acceleration is not as strong when there is a 3-s delay. One way to interpret this finding is that under simultaneous presentation, the relatively greater discriminability for smaller squares is even more apparent and hence has greater influence on dissimilarity judgments.

Figure 7 presents the fit of the range–frequency model of Equation 13 to the mean dissimilarity ratings. The model predictions (solid and dashed lines) appear to capture well the pattern of means (solid and open circles). The squared multiple correlations for the model fits were .98 for the top left panel, .98 for the top right panel, .96 for the bottom right panel, and .97 for the bottom left panel. Although significantly better fits could be obtained by allowing the additive constant a to vary across contexts, the approach taken here was to isolate the effects of context in a single parameter, $1 - w$. The relatively good fit of the range–frequency model of Equation 13 to the dissimilarity ratings lends support to the theoretical assertion that the subjective values being compared in the pairwise dissimilarity task have been altered by the context in a manner consistent with range–frequency theory.

Discussion

The results of Experiment 1 provided strong evidence that increasing the density of contextual stimuli lying between a pair of targets leads to a corresponding increase in perceived dissimilarity. However, the magnitude of this effect depended on task conditions. The constructive–associative model leads to the following interpretations of the data.

First, because contextual effects on pairwise dissimilarity were observed even when squares were simultaneously present and no prior single-stimulus ratings had been made, contextual processing (at least for judgments of squares) is assumed to have fairly high priority. This conclusion must be tempered by the fact that these contextual effects were rather weak as compared with those observed for single-stimulus ratings and for other pairwise rating conditions.

[1] The MDS fits used an algorithm developed by Kruskal (1964). Nearly identical results were obtained using Guttman's (1968) method. The values of stress (following Kruskal's, 1964, Stress Formula 1) were uniformly low for the one-dimensional solutions, varying from .014 to .053, with a mean of .036. Including a second dimension led to an average decrement in stress of only .016. In each case, the second dimension was a horseshoelike configuration on stimulus values. The Shepard (1987) diagrams plotting model distances onto dissimilarities were nonlinear, revealing a strong positive acceleration: Differences in mean ratings near the high end of the scale corresponded to greater differences in model distances than to corresponding differences near the low end of the scale. In addition, the negatively skewed distributions tended to produce greater positive acceleration in the Shepard diagram than the positively skewed distributions under corresponding task conditions. Given that the negatively skewed distributions included more smaller distance comparisons than the positively skewed distribution, this effect is generally consistent with a range–frequency process operating on the distribution of distances.

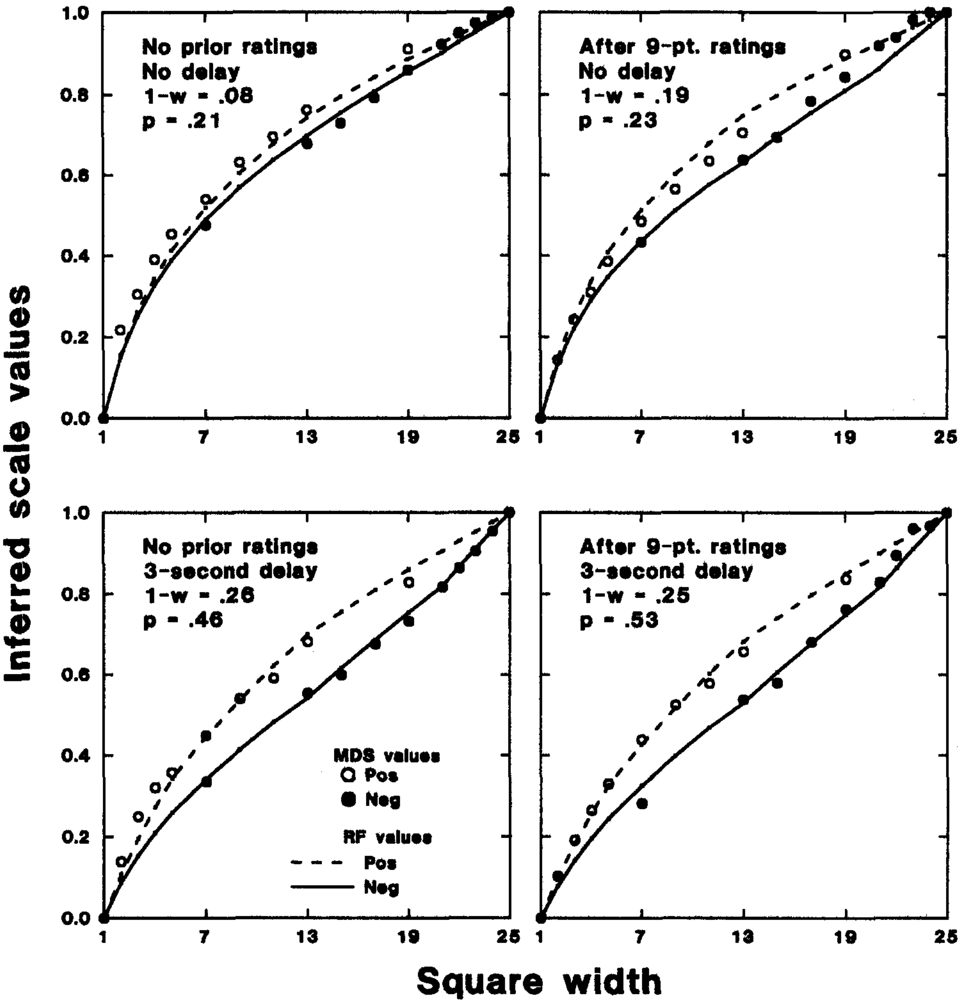

Figure 6. Inferred multidimensional scaling values (filled circles = negative and unfilled circles = positive) and range–frequency derived scale values (solid and dashed lines) from a least squares fit of Equation 13. MDS = multidimensional scaling; RF = range frequency.

Because this finding conflicts with previous research using dot patterns (Mellers & Birnbaum, 1982), some caution must be made in generalizing to other stimulus domains. Experiment 3 examines this issue more closely.

Second, the results provided clear evidence that requiring stimuli to be held in memory for comparison (the 3-s delay conditions) leads to an increase in density effects on similarity. The magnitude of the density effects for these conditions approached that for the single-stimulus ratings. Within the constructive–associative framework, this finding suggests that context-independent scale values may not be easily held in memory, even for simple perceptual stimuli such as squares varying in size. Under delay conditions, it appears that participants use constructive, context-dependent processes to encode stimulus values for later pairwise processing. When one considers the vast array of stimuli for which simultaneous presentation is not a viable possibility (e.g., comparing similarity of personalities, of situations, of motives), this finding suggests that density effects on similarity may be pervasive in real-world judgment situations.

Third, density effects were greater when 9-point ratings of size preceded the similarity task in the 0-s delay condition but were of similar magnitude in the 3-s delay condition. Thus, a set of prior contextually dependent categorical associates may be used to mediate dissimilarity judgments. However, when context-dependent values are being constructed, the establishment of prior associates appears to have little impact.

Finally, the range–frequency model provided a good quantitative description of the single-stimulus ratings, the pairwise dissimilarity ratings, and the MDS values inferred from the pairwise ratings. These results argue for the generality of range–frequency contextual processing to both single-stimulus and pairwise judgment. What did not generalize, however, were the scale values underlying the judgments. For single-stimulus ratings that were not preceded by

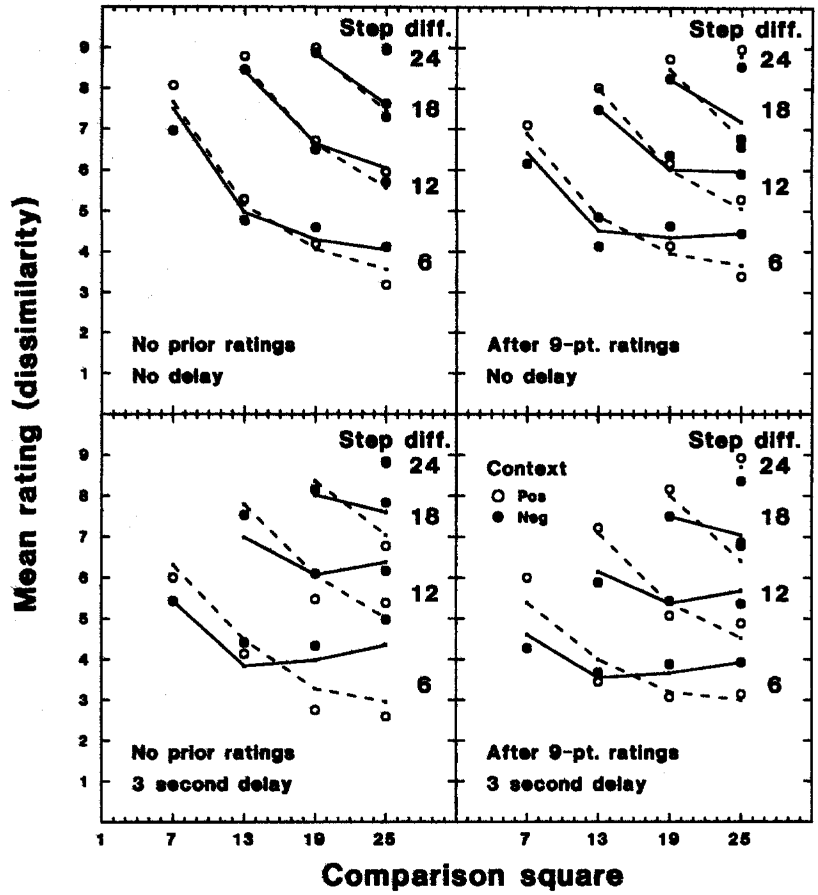

Figure 7. Fit of the range–frequency model of Equation 13 to pairwise dissimilarity ratings of squares in Experiment 1.

similarity ratings, inferred range values were a linear function of square width, replicating prior work (Parducci & Wedell, 1986). However, when pairwise dissimilarity ratings occurred first, the range values for single-stimulus judgment were inferred to be a negatively accelerated power function of stimulus width. The negative acceleration of these functions was even greater for pairwise similarity ratings, especially when squares were simultaneously present. Although all inferred range values were modeled by a power function of square width, the changes in the power exponent demonstrate a lack of scale convergence across tasks for these data.

Experiment 2: Comparing Square Differences Equated on a Power Scale

Because of the differences in the psychophysical scales underlying single-stimulus and pairwise rating tasks demonstrated in Experiment 1, the same-step differences on the single-stimulus scale did not correspond to equal differences on the pairwise scale. As a consequence, there were generally few reversals of rank orders within same-step comparisons. Although disordinal effects were observed and led to different scaling solutions, it would be reassuring if steps were more equally spaced on the psychophysical scale to produce more reversals of rank ordering for same-step comparisons. Experiment 2 replicated the no-prior-rating conditions of Experiment 1, but altered the spacing of the common target stimuli to make equal-step differences more equivalent in the dissimilarity judgment tasks.

Method

The method was the same as the no-prior-rating conditions of Experiment 1 except for the stimuli used. Squares were selected from the set of 25 squares so that they were roughly consistent with a power coefficient of 0.50 for no-delay conditions and 0.67 for delay conditions. Target squares were 1, 4, 9, 16, and 25 for no-delay conditions and 1, 5, 10, 16, and 25 for 3-s delay conditions. Contextual squares were 1.4, 1.8, 2.3, 2.8, 5.4, and 7.1 for positive skew no delay; 11.1, 13.4, 18.8, 20.3, 21.8, and 23.4 for negative skew no delay; 2, 3, 3, 4, 6, and 8 for positive skew 3-s delay; and 12, 14, 17, 19, 21, and 23 for negative skew 3-s delay. Participants were 105 students selected from the same pool as in Experiment 1 and randomly assigned to conditions.

Results and Discussion

The single-stimulus ratings were fit well by the range–frequency model of Equation 12, in which scale values were a power function of square width (R^2s of .997 and .996 between mean ratings and their estimates). For 0-s and 3-s delay conditions, the estimated values of the frequency weighting were 0.27 and 0.28, respectively, and the estimated power exponents were .80 and .69, respectively. These values are in line with those reported earlier for single-stimulus ratings following the dissimilarity rating task (right panel of Figure 4). Once again, the most striking aspect of this result is that the scale values underlying single-stimulus judgment appear to change as a consequence of making prior dissimilarity ratings, becoming negatively accelerated on square width.

The primary goal of Experiment 2 was to replicate the no-prior-rating conditions of Experiment 1, using target stimuli that were spaced more equally along the subjective scale underlying dissimilarity judgments (i.e., negatively accelerated on square width). Figure 8 presents the MDS and range–frequency inferred scales, along with the mean dissimilarity ratings and the range–frequency model fit to those ratings. For both no-delay and delay conditions, the mean rated dissimilarity for same-step comparisons showed the predicted crossover interactions. Because the combined slope for these functions was close to zero, these interactions reflected some reversals of rank ordering. For example, in the delay condition the endpoints of each function show reverse rank ordering across positive and negative skewing conditions. Reversals in rank order were more numerous for the delay condition, but some reversals were still obtained in the no-delay condition. ANOVAs run on mean-rated dissimilarities revealed significant Distribution × Pair interactions for one-step comparisons in the no-delay

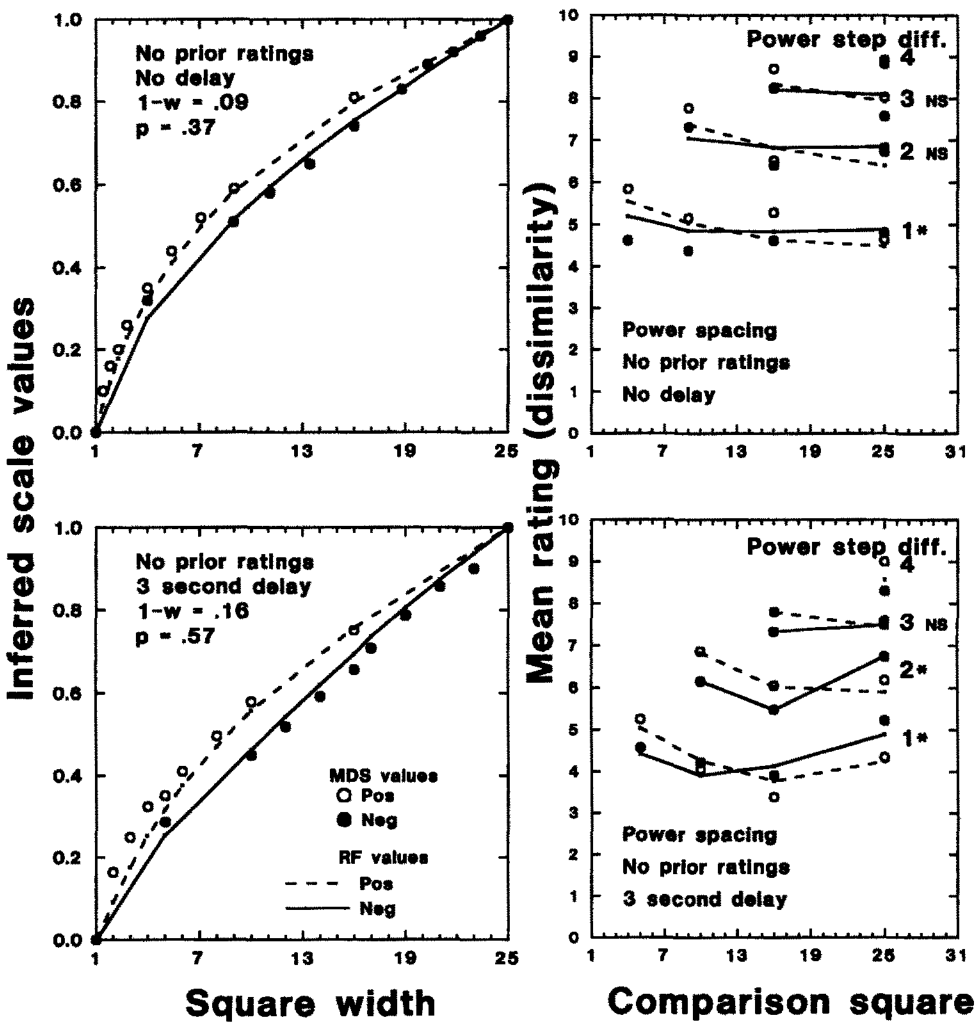

Figure 8. Inferred multidimensional scaling values, range–frequency derived scale values, and range–frequency predictions of pairwise dissimilarity ratings for conditions of Experiment 2. MDS = multidimensional scaling; RF = range frequency.

condition and the one- and two-step comparisons in the delay conditions ($p < .05$).

The MDS recovered scales were again quite similar to the those values inferred by fitting range–frequency theory using Equation 13. The inferred values of $1 - w$ and of the power exponent were similar to those from the corresponding 0-s delay and 3-s delay conditions of Experiment 1. The fit of the range–frequency model to the dissimilarity data captures the basic pattern of data. The squared multiple correlation was .95 for the no-delay data and .97 for the delay data. A significantly better fit would be obtained by allowing the additive constant (a) to vary with context in the no-delay condition. However, the primary purpose of the data fitting was to demonstrate the degree to which context effects on dissimilarities could be handled strictly by changes in $1 - w$.

The results of Experiment 2 provide additional evidence that manipulations of contextual densities can alter the ordering of pairwise dissimilarities even when no prior single-stimulus judgments were made. Although these effects were stronger under the 3-s delay condition, systematic effects of context were also obtained for the 0-s delay condition. The theoretical fits provided in Figures 7 and 8 demonstrate that the contextual effects on mean dissimilarity ratings in Experiments 1 and 2 can be modeled as reflecting underlying shifts in scale values described by range–frequency theory.

Experiment 3: Effects of Prior Ratings and Intrapair Delay on Dissimilarity Judgments of Dot Patterns

Experiment 3 was conducted with two major objectives in mind. The first was to determine the degree to which results from Experiment 1 would generalize to a different psychophysical dimension. The impetus for examining this issue was the apparent discrepancy between results of Experiments 1 and 2 and results reported by Mellers and Birnbaum (1982). Conceptually, their experimental procedure corresponded closely to the no-prior-ratings, 0-s-delay conditions of the present Experiments 1 and 2, but they used dot patterns that varied in number of dots rather than squares that varied in size. Unlike the results reported in the present study for pairwise dissimilarity ratings of squares, Mellers and Birnbaum (1982) found no evidence for density effects on pairwise difference ratings of dot patterns.

There were a number of procedural differences between the Mellers and Birnbaum (1982) method and the method of the present Experiments 1 and 2. These included (a) the dependent variable, which was a rating of the signed difference in the perceived darkness of dot patterns rather than a rating of (unsigned) dissimilarity; (b) the set of stimulus pairs, which included the diagonal entries that paired the stimulus with itself; and (c) the mode of presentation, which consisted of presenting several pairs on the same page of a booklet rather than presenting each pair alone on a computer screen. However, these procedural differences seem fairly trivial from a theoretical perspective.

Less trivial, perhaps, is the difference in psychophysical domains. One explanation why these two domains might differ in sensitivity of pairwise judgments to contextual manipulations arises from analysis of the inferred context invariant scales underlying single and pairwise judgments. In Experiments 1 and 2, the inferred invariant scale values for single-stimulus judgments of squares differed greatly from those for pairwise judgments. The psychophysical scale for single-stimulus judgments was a linear function of square width, but the scale for pairwise judgments was a power function of square width, with the power coefficient close to 0.25 for the relevant conditions. In contrast, Mellers and Birnbaum (1982) demonstrated scale convergence for the two judgment tasks, with the inferred scale values roughly approximated by the logarithm of the number of dots for both single and pairwise judgments. Scale convergence, or lack thereof, may provide an index of the relative fixedness or lability of the psychophysical scale. Viewed in this way, scale values underlying judgments of dot patterns appear to be more fixed, and hence pairwise dissimilarity judgments of dot patterns may be less susceptible to contextual effects than judgments of squares. Experiment 3 explored this possibility by attempting to replicate the null results reported by Mellers and Birnbaum (1982) for pairwise dissimilarity judgments of dot patterns when members of each pair were simultaneously present and no prior single-stimulus judgments have been made. By also varying the delay between pair members and the order of single and pairwise judgment tasks, Experiment 3 examined conditions under which pairwise dissimilarity judgments of dot patterns might be contextually determined.

A second objective of Experiment 3 was to test specific predictions from Krumhansl's (1978) distance–density model. This model describes the modified distance between two stimuli, d'_{ij}, as a function of their interpoint distance in the relevant psychological space, $d(ij)$, and the local densities of contextual stimuli surrounding each stimulus:

$$d'_{ij} = d(ij) + \alpha\delta(i) + \beta\delta(j), \tag{14}$$

where $\delta(i)$ and $\delta(j)$ represent the local densities of stimuli surrounding i and j, respectively, and α and β represent the relative weights of these densities. The signs of the values for α and β are generally assumed to be positive, so that the same interpoint distance corresponds to greater judged dissimilarity when stimuli lie in dense rather than sparse regions.

Experiments 1 and 2 provided strong experimental evidence that increasing densities within a region increases rated dissimilarity of stimuli within that region. These effects of density were well described by Equation 7 and more directly by Equation 13, in which judgments of dissimilarity are an inverse exponential function of values constructed in accordance with principles of range–frequency theory. Indeed, the good fits of the range–frequency model to the dissimilarity ratings and MDS derived values in Experiments 1 and 2 raise the question of whether the local density parameters of Krumhansl's (1978) model are needed at all to explain density effects. That is, if the dimensional locations of the stimuli are determined by a range–frequency process, then the interpoint distance, $d(ij)$, between these locations will reflect the density manipulation without reference to local densities.

However, the local density parameters of Krumhansl's (1978) model become important when one attempts to explain phenomena related to two of the three major objections to geometric models of similarity (Tversky, 1977). These are violations of minimality and symmetry. Tversky pointed out that off diagonal entries of similarity matrices may include values that are higher than diagonal entries, which is an apparent violation of the minimality axiom that the similarity of a point to itself is minimal. Furthermore, Tversky noted that asymmetries are often observed for similarity measures (i.e., the similarity of stimulus i to stimulus j differs systematically from the similarity of j to i. This violates the symmetry axiom according to which $d(ij) = d(ji)$. It is important to note that the constructive–associative framework described by Equations 7 and 8 is not designed to explain these types of density effects, but rather operates within the traditional geometric conception of distance.

These violations of minimality and symmetry are explained within Krumhansl's (1978) model using the concept of local densities. Consider first violations of minimality. If the local densities in Equation 14 receive positive weights, then the perceived dissimilarity of a stimulus from itself will be equal to its interpoint distance, zero, plus positive values for its local density. Thus, the model predicts that the same stimulus will be judged more dissimilar to itself when it occurs within a dense rather than a sparse region. Violations of minimality will occur whenever the sum of the interpoint distance between two stimuli plus their weighted densities is less than the sum of the weighted densities of a single stimulus.

Equation 14 can also be used to predict asymmetries by making assumptions about the relative weights of α and β. Assuming that the similarity of stimulus i to stimulus j is being compared rather than vice versa (either through overt instructions or by presenting i first), Tversky (1977) argued that i would become the focal stimulus and receive greater attention. In Tversky's model, the greater attention translates into greater weight of unique features of the focal stimulus. Within Krumhansl's (1978) model, the greater attention translates into greater weight of the local density surrounding the focal stimulus (i.e., $\alpha > \beta$). Thus, Krumhansl's model predicts that the dissimilarity of i to j will be greater than j to i whenever the density surrounding i is greater than that surrounding j. These predictions of the distance–density model concerning how manipulations of local density affect judged similarity of stimuli to themselves as well as produce asymmetries in dissimilarity judgments were tested in Experiment 3.[2]

Method

Participants and design. The basic 2 × 2 × 2 factorial design was the same as in Experiment 1, with distribution (positive or negative skew), order of tasks (single-stimulus judgments first or pairwise judgments first), and delay between pair members (no delay or 3-s delay) as between-subject variables. The dependent variable for the single-stimulus rating task was the rating of perceived darkness for each of the six target stimuli. The dependent variable for the pairwise rating task was a rating of perceived dissimilarity for the 36 pairings of the six target stimuli. Participants were 196 university students randomly assigned to one of the eight between-subjects conditions, with between 22 and 30 participants per condition.

Apparatus. The microprocessors of Experiment 1 were used to present instructions, display stimuli, and collect responses. Stimuli were displayed in VGA graphics mode as black dots on a white background.

Stimuli. Stimuli were dot patterns that varied in number of dots as described by Mellers and Birnbaum (1982). The six target stimuli consisted of 12, 18, 27, 40, 60, and 90 dots, in accordance with equal logarithmic spacing. Contextual stimuli for the positively skewed conditions consisted of 14, 15, 16, 21, and 23 dots; those for the negatively skewed conditions consisted of 47, 51, 70, 74, and 77 dots. Because there was an additional target stimulus and one less contextual stimulus making up each distribution, the skewing was slightly less than in Experiment 1.

Each dot was roughly circular, with a diameter of 10 pixels. These appeared in black inside white squares that were 200 pixels in width and outlined by a black border. The background on which these squares appeared was also white. On each trial, the dots for each stimulus were randomly assigned to a location within the region outlined by the square border under the constraint that roughly one fourth of the dots appeared in each of the four quadrants within the square. This constraint provided a more uniform distribution of dots than would occur by simple randomization technique and was consistent with the method of Mellers and Birnbaum (1982).

Procedure. The procedure was identical to that used in Experiment 1 except for the following changes. Instructions for single-stimulus presentations asked participants to rate the dot patterns in terms of how light or dark they seemed relative to the other dot patterns using a 9-point scale, ranging from 1 (*very light*) to 9 (*very dark*). Instructions emphasized that dot patterns with few dots should appear light and those with many dots should appear dark. Instructions for pairwise ratings asked participants to rate how similar or dissimilar in darkness the dot patterns appeared to them on a 9-point scale, ranging from 1 (*very very similar*) to 9 (*very very dissimilar*). Again, it was emphasized that dot patterns with few dots should appear subjectively lighter than those with many dots. The only other changes in procedure from Experiment 1 were the presentation sets and orders for the pairwise judgment task. The experimental set of presentations consisted of 121 trials, with each stimulus paired with every other stimulus twice, and once with itself. Order of presentation was randomized for each participant. Which dot pattern appeared on the left of the screen was determined randomly for the first presentation of each pair. On the second presentation of the pair, the arrangement of dot patterns on the screen was reversed.

Results and Discussion

Ratings of darkness. Figure 9 presents the fit of the range–frequency model to the mean ratings of darkness for

[2] In Experiment 1, each stimulus in a pair occurred equally often on the left and right, but because the order of appearance was not recorded with the responses, no tests of asymmetry were possible. Also, because stimuli were not paired with themselves in Experiment 1, tests concerning self-similarity were not possible.

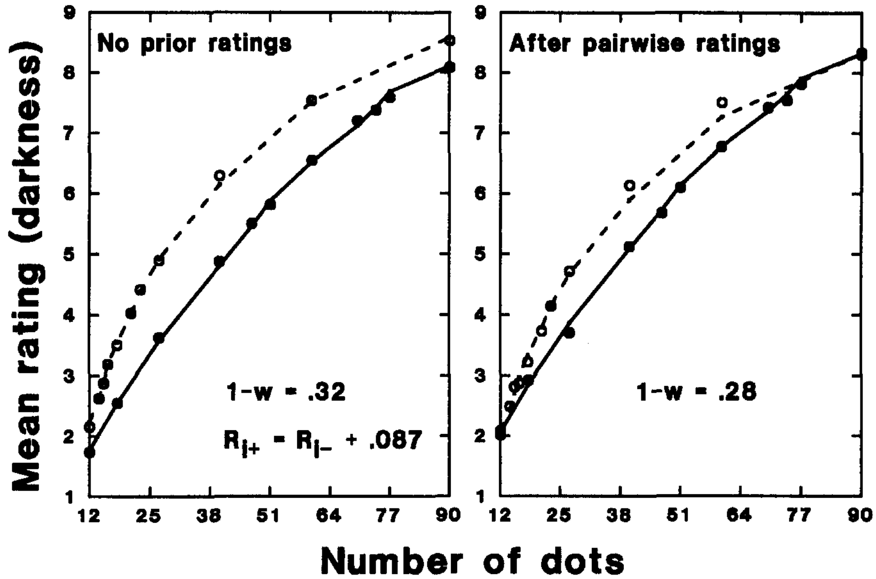

Figure 9. Range–frequency fits to ratings of darkness of dot patterns (Experiment 3).

single-stimulus presentations, combining data from 0-s and 3-s delay conditions.[3] Initial fits equated scale-independent values (S_i) with log (number of dots). However, significant differences between predicted and empirical points led to the selection of a psychophysical function that more accurately predicted ratings. Power and polynomial regression were used to infer the psychophysical function and the best fit was achieved with a second-order polynomial equation: $S_i = 1.884160\phi - 0.009538\phi^2$, where Φ equals the number of dots. Because scale values are on an arbitrary numerical scale, no intercept was fit in the equation. When single-stimulus ratings occurred after pairwise ratings, the model required just three additional fitted parameters, S_{min}, S_{max}, and w. However, when there were no prior dissimilarity ratings, end stimuli were rated higher in the positive-skewing condition than in the negative-skewing condition. This difference was modeled by an additional parameter corresponding to a shift in the range. Thus, for ratings of the left panel, range values for participants in the positive-skewing condition were assumed to shift by 0.097 (on a 0–1 scale), relative to range values for participants in the negative skewing condition. This type of shift is consistent with that reported in previous research when end stimuli are poorly defined and contextual stimuli bunch together near one end stimulus or the other (Parducci et al., 1976).

The fits of the range–frequency model are again quite good using only nine free parameters to model the 44 data points. The value of the frequency weighting parameter is close to that from Experiments 1 and 2. In contrast to the size judgments of squares in Experiment 1, the scale values did not depend on whether pairwise dissimilarity ratings were made prior to single-stimulus ratings. In general, the scale convergence across tasks reported by Mellers and Birnbaum (1982) for these stimuli was replicated, as will be described in further detail below.

A $2 \times 2 \times 2 \times 6$ repeated-measures ANOVA run on the mean ratings of darkness of the six target stimuli confirmed the conclusions described above. A lack of a main effect of, or any interaction with, delay ($p > .10$) justified combining these data in Figure 9. The main effect of target stimulus was highly significant, $F(5, 850) = 2{,}830$, $p < .001$. Unlike Experiment 1, the Target Stimulus $\times$ Judgment Order interaction was nonsignificant ($F < 1$), indicating that the psychophysical scale was unaffected by prior task. The effects of the density manipulation were reflected in the main effect of distribution, $F(1, 170) = 45.2$, $p < .001$, and the Target Stimulus $\times$ Distribution interaction, $F(5, 850) = 21.7$, $p < .001$. In accordance with range–frequency theory, the interaction was primarily contained in the quadratic component, $F(1, 170) = 102.5$, $p < .001$. Finally, a specific contrast comparing ratings of the end-stimuli for the two judgment orders revealed a Judgment Order $\times$ Distribution interaction, $F(1, 170) = 5.3$, $p < .05$. Additional tests (at $p < .05$) demonstrated that the end stimuli were rated significantly higher in the positive skewing condition when there were no prior pairwise ratings, but that distribution had no significant effect on these ratings after the pairwise

[3] Because of some confusion in interpreting instructions, data from 18 participants whose mean ratings of the end stimuli differed by less than one half the range of available categories (i.e., five points) were eliminated. When questioning participants, it was determined that part of confusion appeared to arise from a tendency to rate the luminance of the screen rather than the relative darkness corresponding to the number of dots. Although data for some participants was eliminated for analyses of single-stimulus ratings, none of the data was eliminated for analyses of pairwise ratings. Because the number of dots in the dot patterns constituted the primary way in which they differed, confusion of luminance-based and numerousity-based darkness was apparently not an issue for ratings of dissimilarity.

task. This result supported the additional range-shift parameter used to model data of the left panel of Figure 9.

Ratings of pairwise dissimilarities. Figure 10 presents the mean rated dissimilarity for the 15 pairings of the six target stimuli. If the manipulations of density had no effects on pairwise dissimilarity, the rating functions for each step comparison would be parallel. This appears to be roughly true for the no-delay, no-prior-rating condition (top left panel). Thus, these results generally replicate those of Mellers and Birnbaum (1982) using dissimilarity rather than signed difference judgments.

In stark contrast to the null results of the density manipulation displayed in the upper left panel of Figure 10, mean ratings in the other three panels show systematic effects of density. Consistent with predictions from range–frequency and distance–density models, the slopes for the positively skewed distributions are mostly negative and those for the negatively skewed distribution are mostly positive, indicating systematic shifts in the ordering of pairwise dissimilarities across contextual conditions. The only exceptions to this are the three-step comparisons in the top right panel, in which the reverse ordering is observed.

Separate two-way (Distribution × Target Pair) ANOVAs were conducted on the mean ratings of dissimilarity for the one-, two-, three-, and four-step comparisons of each panel. The significance levels of the interaction terms are presented in Figure 10. As can be seen, when there was no delay between pair members and no prior ratings, none of these interactions achieved statistical significance. However, for each of the other three panels, at least one of these interaction terms was statistically significant. The pattern of results suggests both an associative basis and a constructive basis for these effects. That is, the magnitude of density effects was greater when dissimilarity ratings follow the single-stimulus task, which was designed to establish an associative structure reflecting contextual dependencies. Also, even when no prior associative structure had been formed, the magnitude of density effects was greater when pair members were separated by a 3-s delay than when they were simultaneously present. This implies that a construc-

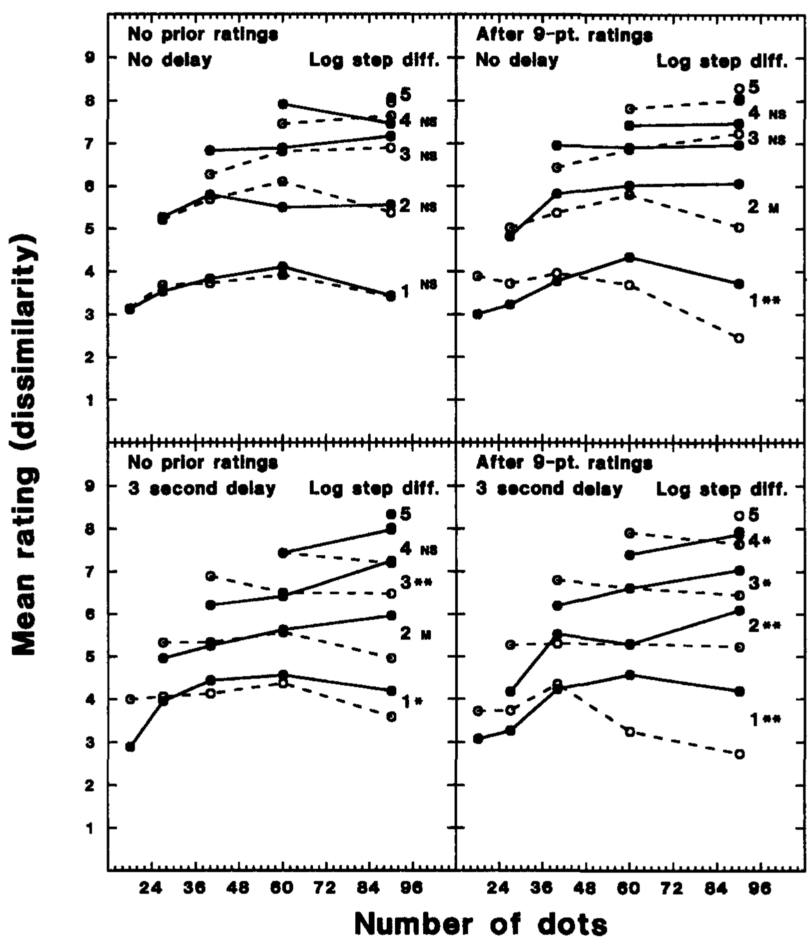

Figure 10. Mean ratings of pairwise dissimilarity for the 15 comparisons among the six target stimuli of Experiment 2. For each pair of functions in each panel, the significance level of the Distribution × Pair interaction term is reported at one of four levels: NS = nonsignificant ($p > .10$), M = marginal ($p < .10$), * = significant ($p < .05$), and ** = significant ($p < .01$).

tive, contextually dependent process is used to encode stimuli when they must be held in memory for even a short period of time.

Results of a 2 × 2 × 2 × 15 repeated measures ANOVA conducted on the mean dissimilarity ratings of the 15 target pairs confirmed this pattern of results. The significant Target Pair × Distribution interaction, $F(14, 2632) = 5.2, p < .001$, reflected the differing effects the distribution manipulation across target pairs. This interaction was moderated both by order of the judgment tasks and delay between pair members as indicated by the significant Target Pair × Distribution × Task Order interaction, $F(14, 2632) = 1.964, p < .05$, and the significant Target Pair × Distribution × Delay interaction, $F(14, 2632) = 2.025, p < .05$. These three-way interactions reflect the greater effects of distribution when prior single-stimulus ratings had been made and when pair members were separated by a 3-s delay. The four-way interaction was not significant, $F < 1.0$.

Range–frequency modeling of dissimilarity data. The method for fitting the range–frequency model to the dissimilarity data differed in one critical aspect than that followed for Experiments 1 and 2. Instead of deriving the psychophysical function from the dissimilarity data and allowing it to vary across the different experimental conditions, the psychophysical function from the single-stimulus rating task was used to predict range–frequency effects on dissimilarity judgments. Thus, following Mellers and Birnbaum's (1982) approach, scale convergence was assumed and a single set of scale values was used to fit data from the single-stimulus and pairwise rating tasks. Once again, the exponential decay function of Equation 11 was used to define the response function. Thus, the following equation was fit to the data using an iterated nonlinear regression method:

$$D_{ij,k} = a - b\exp[-c^*(w \mid R_i - R_j \mid + (1 - w) \mid F_{ik} - F_{jk} \mid)], \quad (15)$$

where R_i and R_j were taken from the fit to the single-judgment data, but rescaled so that range values for the fewest and most dots were 0 and 1, respectively. Altogether, only four parameters are free to vary in Equation 15, and only one of these, $1 - w$, is used to explain contextual effects.

Figure 11 compares the inferred scale values from fitting Equation 15 to the dissimilarity data with the nonmetric MDS values. Again, one-dimensional solutions were selected for the MDS values on the basis of a priori considerations as well as the low values of stress for these solutions.[4] The correspondence between range–frequency-inferred scale values and the MDS values is impressive, especially when one considers that the functions in each panel were generated from a single-contextual parameter $(1 - w)$ and range values taken from the single-stimulus judgment task. The values of $1 - w$ inferred for each condition are consistent with the pattern of results described earlier. When there was no delay and no prior ratings, there was virtually no effect of manipulating stimulus densities. However, prior single-stimulus ratings provided contextually altered categorical associates that resulted in context effects on the inferred scale values underlying pairwise judgment. Interspersing a 3-s delay between presentation of the members of each pair led to strong context effects of similar magnitude as found for squares. According to the constructive–associative framework, this increase in context effects implies that context-independent scale values are not easily held in memory. Instead, memory constraints appear to initiate a process whereby context-dependent stimulus values are constructed and subsequently used as a basis for dissimilarity judgments.

Figure 12 presents the fit of the range–frequency model of Equation 15 to the dissimilarity data. The squared multiple correlations for the four conditions were .98 for the top left panel, .96 for the top right, .98 for the bottom left, and .97 for the bottom right. Although several of these deviations from predicted values are rather large, the model generally captures the differences in ratings associated with context.

Tests of the distance–density model. Figure 13 presents mean ratings of dissimilarity for target stimuli paired with themselves under the four basic task conditions of Experiment 3. Assuming positive weights for local densities, Krumhansl's (1978) distance–density model (Equation 14) predicts that the perceived dissimilarity of a stimulus from itself should increase when it lies in a dense rather than a sparse region of the psychological space. Thus, Krumhansl's model predicts a crossover interaction of the form that the zero-step functions of Figure 13 should have a negative slope for positively skewed distributions and a positive slope for negatively skewed distributions. This prediction is consistent with the pattern of data for three of the four task conditions, the exception being the no-delay, no-prior-ratings condition for which density effects were minimal.

A 2 × 2 × 2 × 6 repeated measures ANOVA was conducted on the self-dissimilarity ratings of these six target stimuli. The predicted Target × Distribution interaction was statistically significant, $F(5, 945) = 2.5, p < .05$. The higher order interactions involving distribution did not achieve statistical significance ($p > .10$). Given the density manipulations of Experiment 3, Krumhansl's (1978) model predicts that the strongest effects of density on self-dissimilarity would be for Targets 18 and 60. An examination of the means showed this to be true (averaging across task conditions, $M_{18} = 3.04$ and $M_{60} = 2.62$ for positive skewing; $M_{18} = 2.30$ and $M_{60} = 2.98$ for negative skew-

[4] The MDS fits followed the same procedure as described in Experiment 1 (see Footnote 3). The values of stress (following Kruskal's, 1964, Stress Formula 1) were slightly higher than in Experiment 1 but were still sufficiently low for the one dimensional solutions, varying from .020 to .079, with a mean of .050. Including a second dimension led to an average decrement in stress of only .018, and as in Experiment 1 the second dimension was a horseshoelike configuration on stimulus values. Unlike Experiment 1, the Shepard (1987) diagrams plotting model distances onto dissimilarities were nearly linear for both positively and negatively skewed sets.

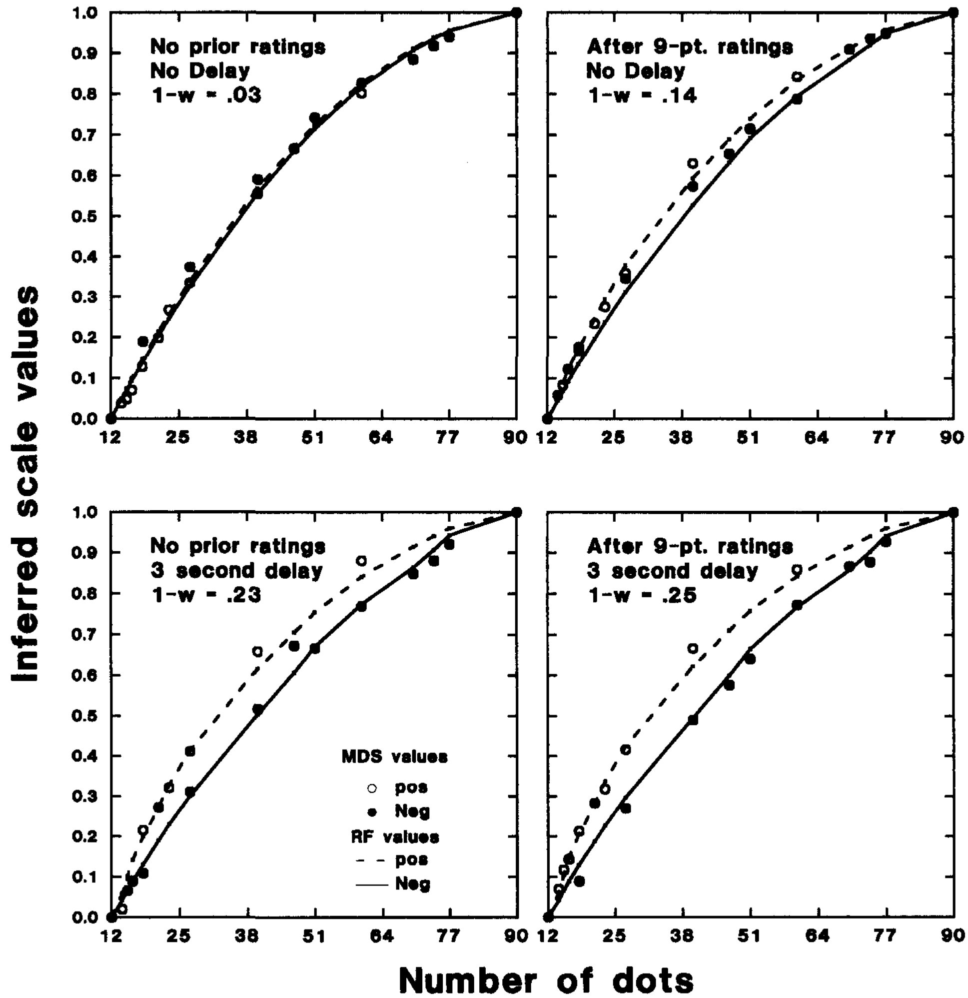

Figure 11. Inferred multidimensional scaling values (solid and open circles) and range–frequency derived scale values (lines) from least squares fit of Equation 15. MDS = multidimensional scaling; RF = range frequency.

ing). A 2 × 2 × 2 × 2 repeated measures ANOVA conducted on the ratings of these two targets confirmed the strength of this Target × Distribution interaction, $F(1, 189) = 10.1, p < .01$. In addition, the Target × Distribution × Judgment Order interaction was statistically significant, $F(1, 189) = 6.1, p < .05$, but none of the other interaction terms involving distribution were significant. This three-way interaction suggests that prior categorization may be at least partially responsible for the density effects on self-dissimilarity. In line with this assertion, separate 2 × 2 ANOVAs conducted on the mean ratings of Targets 18 and 60 for each of the four task conditions revealed that the Target × Distribution interaction terms were statistically significant whenever single-stimulus ratings preceded judgment ($p < .05$), but were nonsignificant whenever pairwise judgments occurred first ($Fs < 1.0$).

A second prediction of the distance–density model tested in Experiment 3 concerned density-induced asymmetries in dissimilarity judgments. Under the assumption that the weight of the local densities of the first stimulus presented is greater than that for the second stimulus (i.e., $\alpha > \beta$), the model predicts that a stimulus pair will be judged more dissimilar when the pair member with the higher local density is presented first. To test this prediction, effects of presentation order were examined for nine pairs of target stimuli in the delay conditions. The nine pairs, shown in Figure 14, were selected so that one member of the pair was located in a sparse region and one in a dense region of the psychological space. As shown in Figure 14, the effects of order of presentation on mean ratings of dissimilarity are generally consistent with the predictions of the distance–density model. For the positively skewed distribution, presenting the lower valued stimulus first (i.e., the stimulus from the denser region) resulted in higher ratings of dissimilarity. For the negatively skewed distribution, this order effect was strongly reduced and reversed for some target pairs.

Initially, a 2 × 2 × 2 × 2 × 9 repeated measures ANOVA was conducted on the mean ratings of similarity

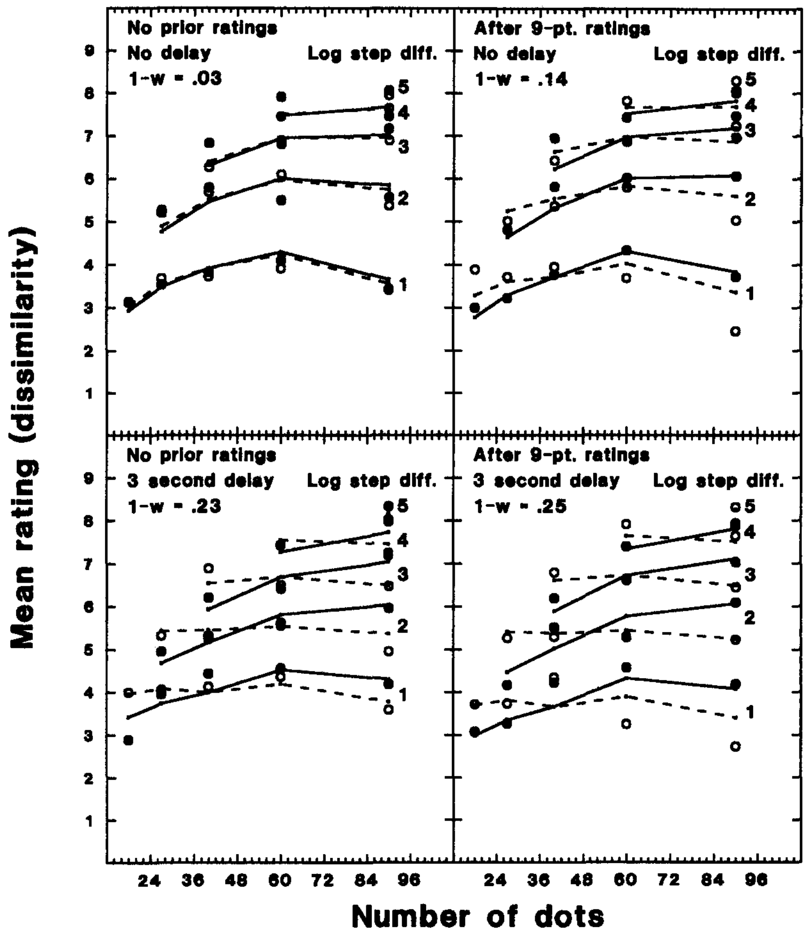

Figure 12. Fit of the range–frequency model of Equation 15 to pairwise dissimilarity ratings of dot patterns in Experiment 4. Note that the scale values are derived from the fit to single-stimulus judgments of Figure 11.

for the nine target pairs under the two order conditions for the entire set of participants. The predicted asymmetry effect was reflected in a significant Pair Order × Distribution interaction, $F(1, 189) = 5.8$, $p < .05$. However, a significant Pair Order × Distribution × Delay interaction, $F(1, 189) = 6.6$, $p < .05$, indicated that the form of the two-way interaction depended on amount of delay. Separate ANOVAs conducted for 0-s and 3-s delay conditions revealed that the Pair Order × Distribution interaction was not significant for the no-delay conditions ($F < 1$), but was significant for the 3-s delay conditions shown in Figure 14, $F(1, 99) = 10.9, p < .01$. It is not particularly surprising that asymmetry effects did not occur for simultaneous presentation, because in this condition there is no particular reason to designate one of the pair members a focal stimulus. Further analysis of the 3-s delay data revealed a significant main effect of pair order, $F(1, 99) = 7.6, p < .01$, reflecting the overall tendency to rate a pair more dissimilar when the pattern with fewer dots was presented first. This result is consistent with the scale values inferred from the range–frequency modeling of the psychophysical function, in which Targets 12, 18, and 27 were more closely spaced together than were Targets 40, 60, and 90. Finally, the lack of a Pair Order × Distribution × Task Order interaction ($p > .10$) suggests that the asymmetry effects were of similar magnitude regardless of task order and thus justifies combining the two orders in Figure 14.

Comparison of the distance–density and range–frequency models. Because the effects of density on self-similarity and direction of comparison are predicted by the distance–density model but not by the range–frequency model, one may ask whether the distance–density model provides an adequate account of the full pattern of data? To fit the density model, one needs to estimate three parameters, in addition to those defining the psychophysical function and response function. These are the weighting parameters α and β, and also the radius or length of interval defining local density. Because density is the frequency of stimuli falling within a fixed radius, when the radius is zero, there will be no density effects. As the radius increases, density effects increase at first, but when the radius includes the full set of stimuli, density effects are eliminated.

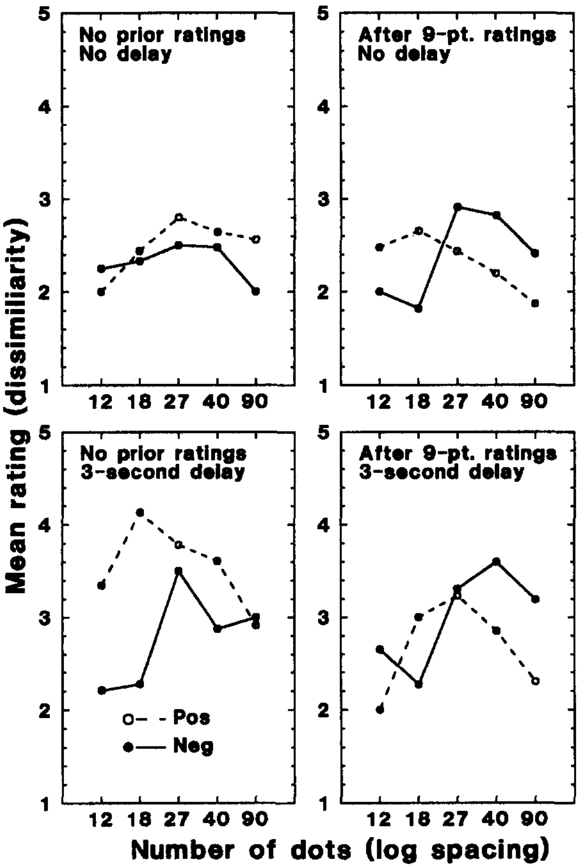

Figure 13. Density effects on ratings of dissimilarity of target stimuli to themselves. The Stimulus × Distribution interaction is consistent with predictions from Krumhansl's (1978) distance–density model.

There appear to be two problems in using the distance–density model as a sole explanation of the full pattern of density effects. The first of these arises from the relationship between density effects on pairs of identical stimuli and density effects on pairs of nonidentical stimuli, the latter of which constitute the focus of the range–frequency modeling. Within the distance–density model, both effects will occur whenever the density weights (α and β) are greater than zero. Thus, if no density effects are observed for one type of comparison, they cannot occur for the other type of comparison. In the no-prior-ratings 3-s-delay condition of Experiment 3, the density effects on pairs of identical stimuli were not significant, and yet significant density effects on pairs of nonidentical stimuli did occur. This pattern of effects is inconsistent with the distance–density model. However, one may argue that the nonsignificant density effects on self-similarity may result from a lack of statistical power and therefore do not necessarily contradict the distance–density model.

A second, more serious problem is that the distance–density model predicts an effect in the opposite direction than was observed. The predictions concern the 18-step comparisons of Experiment 1, the three-power-step comparisons of Experiment 2, and the four-log-step comparisons of Experiment 3. In each of these comparisons, an end stimulus is compared with the target stimulus closest to the other end stimulus. For example, in Experiment 1 the 18-step comparisons corresponded to comparisons of Squares 1 to 19 and 7 to 25. The anomalous prediction arises from the fact that the end stimuli tend to lie in a less dense region than the closest corresponding target stimulus. For example, the density surrounding Square 1 in the positively skewed distribution includes stimuli between Squares 1 and 7, but the density surrounding Square 7 includes both squares between Squares 1 and 7 and between Squares 7 and 13. Thus, the distance–density model predicts greater dissimilarity for the 7–25 comparison than the 1–19 comparison in the positive-skewing condition and the reverse pattern in the negative-skewing condition. This is the opposite prediction of the range–frequency model. An examination of the data

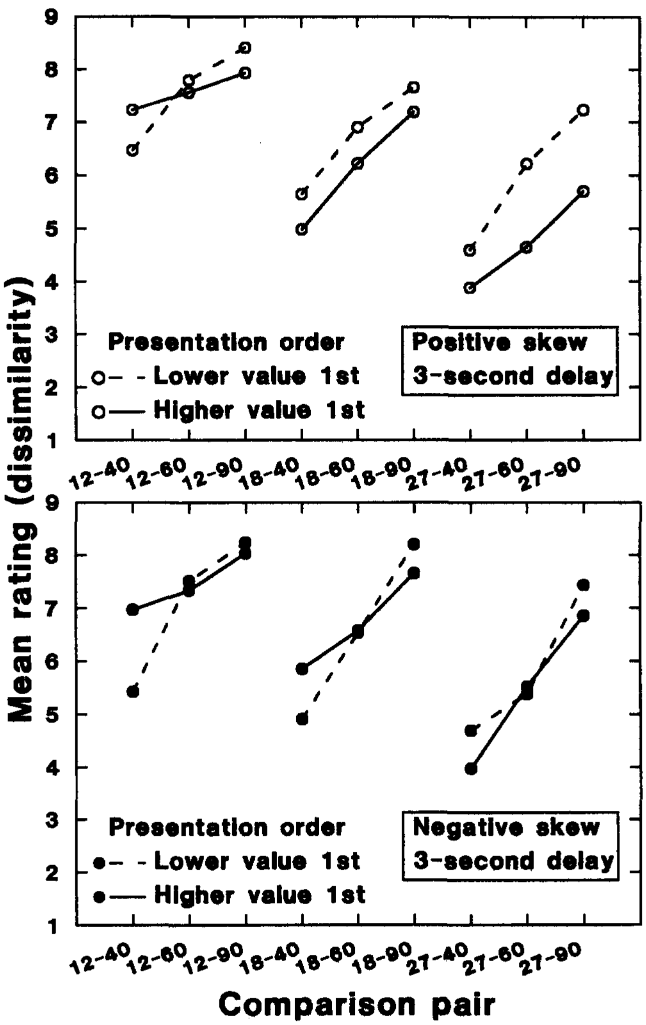

Figure 14. Effects of order of comparison on dissimilarity judgments for positive and negative skewing, 3-s delay conditions. The Order × Distribution interaction is consistent with predictions from Krumhansl's (1978) distance–density model.

reveals that the two significant interactions found for the relevant comparisons were both in the direction predicted by the range–frequency model and inconsistent with the distance–density model.

Finally, the distance–density and range–frequency models can be directly compared by assessing how well they fit the data of Experiment 3. The data points fit by the models corresponded to those shown in Figure 10, which averaged across direction of comparison and excluded pairs of identical stimuli. The fit of the range–frequency model is already shown in Figure 12. To fit the distance–density model, the same psychophysical function derived from the single-stimulus rating task was used. As with the range–frequency fit, the fitting procedure placed the parameters of the distance–density model of Equation 14 into the inverse exponential response function of Equation 11. Because asymmetries were not being modeled, the two weighting parameters, α and β, were constrained to equal one another. Like the range–frequency model, the distance–density model fit the data for each Delay $\times$ Prior-Response-Task condition with three response parameters and a weighting parameter that modeled context effects. The distance–density model required one additional parameter corresponding to the radius defining local density fit to each condition.

A hierarchical regression of the 120 data points of Figure 10 on the two sets of model estimates was conducted to compare the model fits. When the range–frequency model estimates were entered first, the change in the squared multiple correlation corresponding to inclusion of the distance–density model estimates was nonsignificant, $F(1, 117) = 2.016$, $p > .05$. When the distance–density model estimates were entered first, the change in the squared multiple correlation corresponding to inclusion of the range–frequency model estimates was significant, $F(1, 117) = 22.5$, $p < .001$. Thus, using identical fitting procedures and fewer fitted parameters, the range–frequency model estimates accounted for significantly more variance in the mean dissimilarity ratings than did the distance–density model predictions.

General Discussion

The experimental results reported here provide clear and compelling evidence that increasing contextual densities within a region can lead to an increase in the perceived dissimilarity of stimuli within that region. Previous research in this area has either applied density principles in a post hoc manner to explain effects of nonmanipulated variables (Appleman & Mayzner, 1982; Krumhansl, 1978; Nosofsky, 1991) or has found null (or small assimilative) effects of density manipulations (Corter, 1987; Mellers & Birnbaum, 1982). Thus, the present results provide a solid empirical basis for theoretical assertions that contextual densities can affect similarity.

Constructive–Associative Framework

The constructive–associative model of contextual effects described in the present article provides a framework for understanding when contextual densities should affect similarity as well as what types of mechanisms guide these effects. By far, the strongest density effects in Experiments 1–3 occurred when the stimuli being compared had to be held in memory for a short time. The fact that density effects were greatly reduced or eliminated when memory constraints were removed implies that the processes for representing the stimuli in memory were context dependent. These processes are constructive in the sense that they do not require the prior establishment of a context-dependent associative structure. Thus, one way to interpret the effects of the delay manipulation is that under memory constraints, range–frequency processes provided a context-dependent encoding of the stimuli that was then used as a basis for judging similarity.

It is perhaps not too surprising that density effects are greatly reduced or eliminated when participants judge the dissimilarity of a pair of psychophysical stimuli that are presented side by side. In this situation, there seems little reason to compare the targeted stimuli with contextual values residing in memory to establish values for these stimuli. After all, the physical values for these stimuli are directly available. Furthermore, the task is concerned with differences between the stimuli and, hence, the stimulus could be conceived as the result of a subtractive process operating on the physical codes directly. Under this conceptualization, the context becomes the distribution of pairwise differences rather than the stimulus distributions themselves. It seems all the more remarkable then that constructive processes appeared to operate on squares presented simultaneously with no prior single-stimulus ratings. Apparently, even psychophysical domains can differ in the priority of contextual processing, and as speculated earlier, this priority may be related to the lability or stability of the underlying psychophysical function. More generally, these results suggest that density effects should be more pervasive in domains for which direct comparison of simultaneously presented stimuli is not feasible.

The results of Experiment 3 also provided compelling evidence for an associative basis of density effects on pairwise dissimilarity. When targets were presented simultaneously with no prior rating task, there were no effects of context on dissimilarity ratings. However, when the pairwise task was preceded by single-stimulus ratings, significant density effects on pairwise dissimilarity emerged. These effects can be interpreted as resulting from the context-dependent associative network established during the prior task. Thus, even when pair members were presented simultaneously so that direct comparison of physical values was possible, retrieved categorical associates appeared to influence perceived dissimilarity. For small-step comparisons, increased density in a region led to increased dissimilarity, as would be expected if dissimilarity were based on the difference in categorical encodings from the single-stimulus rating task. Goldstone (1994) has recently demonstrated increased discrimination for stimuli along a physical dimension that is being used as a basis for categorization. This type of enhanced discriminability between stimuli assigned to different categories may represent the same type of associative process described here for increasing dissimilarity on the basis of manipulation of prior category ratings.

Processes Underlying Density Effects

Although the constructive–associative framework as described by Equations 1–8 provided a good explanation of most of the density effects reported in this article, it does not explain effects of density on self-similarity and asymmetry of judgments reported in Experiment 3. This is because the associative–constructive framework operates within the traditional axioms of geometric models of similarity. These effects were generally consistent with predictions from Krumhansl's (1978) distance–density model (Equation 14), which incorporates the idea of local densities operating on judgments. Thus, these results provide strong experimental evidence that densities are predictive of phenomena that appear to violate the axioms underlying geometric representations of similarity. Furthermore, an alternative explanation of these effects in terms of Tversky's (1977) feature-based model of similarity seems implausible, because the features of the stimuli did not differ across density manipulations. There is an additional problem in applying Tversky's model (1977) to explain the results. Nosofsky (1991) has pointed out that according to Tversky's model (1977), the asymmetric relationship in which the similarity of i to j is greater than the similarity of j to i implies that the self similarity of j to j will be greater than the self-similarity of i to i. However, the opposite relationship was obtained in Experiment 3. These results are consistent with Krumhansl's (1978) model and inconsistent with Tversky's model (1977).

One criticism of Krumhansl's (1978) theoretical approach is that there is little in the way of a psychological mechanism specified to explain density effects. Why should there be local context effects tied to the densities of surrounding stimuli? One possibility is suggested by the context-evoking properties of stimuli as described in Kahneman and Miller's (1986) norm theory. According to norm theory, a stimulus often evokes similar stimuli for comparison. As applied to density effects on pairwise dissimilarity judgments, presentation of a stimulus may result in automatic retrieval of nearest neighbors. The distributional set of nearest neighbors evoked would depend on the contextual set of stimuli. Stimuli closer in value to the stimulus are more likely to be evoked when that stimulus is located in a dense rather than in a sparse region. The distances between these evoked values and the stimulus may then provide a norm for judging similarity to the comparison stimulus. Assuming some errors in the discrimination process, a stimulus would then be judged less similar to itself when it occurs in a dense rather than in a sparse region because its distance from itself is larger relative to the distances of nearest neighbors. Asymmetry effects would also result from this process because the same interpoint distance would seem large compared with the small distances evoked by the nearest neighbors in a dense region, and would seem small when compared with the large distances evoked by the nearest neighbors in a sparse region.

Other explanations of density effects have been described in the literature. Ennis (1988) has argued that density-related effects can be modeled in terms of variability associated with the location of a stimulus in the multidimensional space. Ashby and Perrin (1988) have applied a signal detection analysis to argue that judgment criteria may differ for different pairs, depending on their relative frequency. The present data were not aimed at distinguishing between these different interpretations of local density effects, but rather on establishing the existence of such effects.

Indeed, the focus of the present experiments was on global effects of density on dissimilarity judgments more than local effects associated with violations of the minimality and symmetry axioms. The model fitting of Experiments 1–3 indicate that a reasonable account of these global effects of density is given by assuming that dissimilarity is a function of the difference in values resulting from a range–frequency judgment process. In Experiment 3, the range–frequency model provided a significantly better fit of global density effects than did the distance–density model, which used more fitted parameters. Furthermore, the fact that significant density effects on nonidentical pairs were obtained when there were no significant effects on self-similarity calls into question whether these effects result from the same process. As Nosofsky (1991) has pointed out, density effects may operate at both the stimulus and response level. Because the self-similarity effects depended on prior categorical encoding in Experiment 3, these effects may be more reflective of a response-based process.

Geometric and Feature-Based Representations of Similarity

What are the implications of this research for the theoretical conceptualization of how similarity is represented? For example, should we conceive of similarity in terms of geometric or a feature-based representations? This question seems somewhat misguided in that it implies that one perspective excludes the other. Tversky and his colleagues have provided strong evidence that manipulations of common and distinctive features produce effects on similarity that are generally in line with his feature-based model (Gati & Tversky, 1984; Tversky, 1977; Tversky & Gati, 1982). However, it is difficult to see how a feature-based representation would account for the strong density effects reported here because the features of the contextual stimuli were held constant across density manipulations. It seems more reasonable to assume that people are flexible processors of information and can generate a variety of representations that may serve as bases for similarity.

Consistent with this flexibility is the finding that within both feature-based and geometric representations, similarity is not invariant but rather highly dependent on the stimulus context. Roberts and Wedell (1994) demonstrated significant effects of context on the inferred similarity between emotion words when a sorting task was used but not when a direct pairwise similarity rating task was used. These contextual effects were modeled within the geometric framework as the emergence of a third dimension in different contextual conditions. Other researchers have consistently demonstrated contextual effects on similarity modeled within a feature-based framework (Tversky, 1977; Tversky & Gati, 1982). The flexibility of similarity relations may be a cause for concern and may even be viewed as rendering the construct theoretically useless as an explanatory construct (Goodman, 1972). However, as Medin, Goldstone, and Gentner (1993) have argued, an understanding of how similarity relations relate to features of the stimuli and the task can result in a more complete understanding of similarity that ultimately may prove useful as an explanatory construct.

References

Appleman, I. B., & Mayzner, M. S. (1982). Application of geometric models of letter recognition: Distance and density. *Journal of Experimental Psychology: General, 111*, 60–100.

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

Beebe-Center, J. G. (1929). The law of affective equilibrium. *American Journal of Psychology, 41*, 54–69.

Birnbaum, M. H. (1974). Using contextual effects to derive psychophysical scales. *Perception & Psychophysics, 15*, 89–96.

Corter, J. E. (1987). Similarity, confusability, and the density hypothesis. *Journal of Experimental Psychology: General, 116*, 238–249.

Corter, J. E. (1988). Testing the density hypothesis: Reply to Krumhansl. *Journal of Experimental Psychology: General, 117*, 105–106.

DeSarbo, W. S., Manrai, A. K., & Burke, R. R. (1990). A nonspatial methodology for the analysis of two-way proximity data incorporating the distance–density hypothesis. *Psychometrika, 55*, 229–253.

Di Lollo, V., & Casseday, J. H. (1965). Graded contrast effects in the judgment of lifted weights. *Journal of Experimental Psychology, 70*, 234–235.

Ennis, D. M. (1988). A multidimensional stochastic theory of similarity. *Journal of Mathematical Psychology, 32*, 449–465.

Gati, I., & Tversky, A. (1984). Weighting common and distinctive features in perceptual and conceptual judgments. *Cognitive Psychology, 16*, 341–370.

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123*, 178–200.

Goodman, N. (1972). Seven strictures on similarity. In N. Goodman (Ed.), *Problems and projects* (pp. 437–447). New York: Bobbs-Merrill.

Guttman, L. A. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika, 33*, 469–506.

Haubensak, (1992). A process model for absolute judgments. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 303–309.

Helson, H. (1947). Adaptation-level as frame of reference for prediction of psychophysical data. *American Journal of Psychology, 60*, 1–29.

Hollingsworth, H. L. (1910). The central tendency of judgment. *Journal of Philosophy, Psychology, and Scientific Method, 7*, 461–468.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*, 411–428.

Hunt, W. A., & Volkmann, J. (1937). The anchoring of an affective scale. *American Journal of Psychology, 49*, 88–92.

Johnson, D. M. (1944). Generalization of a scale of values by the averaging of practice effects. *Journal of Experimental Psychology, 34*, 425–436.

Johnson, D. M. (1949). Learning function for a change in the scale of judgment. *Journal of Experimental Psychology, 39*, 316–321.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*, 136–153.

Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review, 85*, 445–463.

Krumhansl, C. L. (1982). Density versus feature weights as predictors of visual identification: Comments on Appleman and Mayzner. *Journal of Experimental Psychology: General, 111*, 101–108.

Krumhansl, C. L. (1988). Testing the density hypothesis: Comment on Corter. *Journal of Experimental Psychology: General, 117*, 101–104.

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika, 29*, 115–129.

Medin, D. L., Goldstone, R., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*, 254–278.

Medin, D. L., & Heit, E. (in press). Categorization. In D. E. Rumelhart & B. O. Martin (Eds.), *Handbook of cognition and perception: Cognitive science*. San Diego, CA: Academic Press.

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

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology, 35*, 113–118.

Mellers, B. A. (1983). Equity judgment: A revision of Aristotelian views. *Journal of Experimental Psychology: General, 111*, 242–270.

Mellers, B. A., & Birnbaum, M. H. (1982). Loci of contextual effects in judgment. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 582–601.

Mellers, B. A., & Birnbaum, M. H. (1983). Contextual effects in social judgment. *Journal of Experimental Social Psychology, 19*, 157–171.

Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology, 23*, 94–140.

Nosofsky, R. M. (1992). Exemplar-based approach to relating categorization, identification, and recognition. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 363–393). Hillsdale, NJ: Erlbaum.

Parducci, A. (1956). Direction of shift in the judgment of single stimuli. *Journal of Experimental Psychology, 51*, 169–178.

Parducci, A. (1963). Range–frequency compromise in judgment. *Psychological Monographs, 77* (92, Whole No. 565).

Parducci, A. (1965). Category judgment: A range–frequency model. *Psychological Review, 72*, 407–418.

Parducci, A. (1982). Scale values and phenomenal experience: There is no psychophysical law! In H.-G. Geissler & P. Petzold (Eds.), *Psychophysical judgment and the process of perception* (pp. 11–16). Amsterdam: North-Holland.

Parducci, A. (1983). Category ratings and the relational character of judgment. In H. G. Geissler, H. F. J. M. Buffort, E. L. J. Leeuwenberg, & V. Sarris (Eds.), *Modern issues in perception* (pp. 89–105). East Berlin, German Democratic Republic: Deutscher Verlag der Wissenschaften.

Parducci, A., Knoble, S., & Thomas, C. (1976). Independent contexts for category ratings: A range–frequency analysis. *Perception & Psychophysics, 20,* 360–366.

Parducci, A., & Perrett, L. F. (1971). Category rating scales: Effects of relative spacing and frequency. *Journal of Experimental Psychology Monograph, 89,* 427–452.

Parducci, A., & Wedell, D. H. (1986). The category effect with rating scales: Number of categories, number of stimuli, and method of presentation. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 496–516.

Riskey, D. R., Parducci, A., & Beauchamp, G. K. (1979). Effects of context in judgments of sweetness and pleasantness. *Perception & Psychophysics, 26,* 171–176.

Roberts, J. S., & Wedell, D. H. (1994). Context effects of similarity judgments of multidimensional stimuli: Inferring the structure of the emotion space. *Journal of Experimental Social Psychology, 30,* 1–38.

Shepard, R. N. (1987, September 22). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323.

Smith, R. H., Diener, E., & Wedell, D. H. (1989). Intrapersonal and social comparison determinants of happiness: A range–frequency analysis. *Journal of Personality and Social Psychology, 56,* 317–325.

Thurstone, L. L. (1927). Psychophysical analysis. *American Journal of Psychology, 38,* 368–389.

Tresselt, M. E. (1947). The influence of amount of practice upon the formation of a new scale of judgment. *Journal of Experimental Psychology, 69,* 241–245.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352.

Tversky, A., & Gati, I. (1982). Similarity, separability, and the triangle inequality. *Psychological Review, 89,* 123–154.

Wedell, D. H. (1984). *A process model for psychophysical judgment.* Unpublished doctoral dissertation, University of California, Los Angeles.

Wedell, D. H. (1994). Contextual contrast in evaluative judgments: A test of pre- versus postintegration models of contrast. *Journal of Personality and Social Psychology, 66,* 1007–1019.

Wedell, D. H., & Parducci, A. (1988). The category effect in social judgment: Experimental ratings of happiness. *Journal of Personality and Social Psychology, 55,* 341–356.

Wedell, D. H., Parducci, A., & Lane, M. (1990). Reducing the dependence of clinical judgment on the immediate context: Effects of number of categories and type of anchors. *Journal of Personality and Social Psychology, 58,* 319–329.

Wedell, D. H., Parducci, A., & Roman, D. (1989). Student perceptions of fair grading: A range–frequency analysis. *American Journal of Psychology, 102,* 233–248.

Wilkinson, L. (1988). *SYSTAT: The system for statistics.* Evanston, IL: SYSTAT.

Received August 5, 1994
Revision received March 13, 1995
Accepted April 13, 1995 ■