Contrast Effects in Paired Comparisons: Evidence for Both Stimulus-Based and Response-Based Processes

Douglas H. Wedell
University of South Carolina

Participants made paired comparisons of square sizes, with blue squares presented on the left side of a computer screen and red squares on the right. Context was manipulated by varying the distributions of blue and red squares separately. In Experiment 1, target squares from low-range distributions were judged larger than the same-size squares from high-range distributions. In Experiment 2, ranges were equated but ranks manipulated between distributions, pairing bell with U-shaped and positively skewed with negatively skewed distributions. Results provided little support for a rank-dependent valuation model. Experiment 3 used distributions that were designed to test between adaptation-level and response-equalization models. Both models received support, with response latencies constituting an important moderating variable. Response patterns for short latency participants were consistent with a stimulus-based adaptation-level process, and those for longer latency participants were consistent with a response-equalization process.

In the paired-comparisons procedure, participants indicate which member of a presented pair is greater in value on a specified attribute dimension. The method of paired comparisons has a long history within psychology, providing the basis for early psychophysical laws described by Fechner (1860/1966) and for fundamental scaling procedures still commonly used today (Davidson, Suppes, & Seigel, 1957; Luce, 1959; Thurstone, 1927a, 1927b). One appealing facet of the method is that it appears to provide a clear basis for ordering stimuli along a dimension. If the probability of choosing stimulus X over stimulus Y significantly exceeds .50, then X may be assumed to dominate Y on the relevant psychological dimension.[1]

However, invariance of stimulus ordering is not always obtained with paired comparisons. For example, Tversky (1969) demonstrated intransitivities of pairwise choice using multidimensional stimuli. Disordinal effects with multidimensional stimuli can be explained in terms of differential weighting of attributes, differential valuing of attributes, or use of heuristic strategies such as valuing dominance in relation to other stimuli (Wedell, 1991). These explanations do not apply to context effects on the ordering of unidimensional stimuli, which have been demonstrated predominantly with psychophysical materials. For example, early work by Helson and his colleagues (Helson, 1964; Helson, Michels, & Sturgeon, 1954) demonstrated that the point of subjective equality (PSE) of a stimulus derived from paired-comparisons procedures could be systematically varied by manipulating the contextual set of stimuli to which the standard was compared. The same comparison stimulus was

assigned a lower value relative to the standard when the distribution of stimuli consisted of mostly higher valued stimuli. This type of contrast effect in paired comparisons meant that the inferred ordering of stimuli relative to the standard depended on the set of stimuli used for comparison. Contrast effects in paired comparisons have been demonstrated to generalize across a wide variety of stimulus dimensions, including heaviness of lifted weights (Helson et al., 1954; Parducci & Haugen, 1967; Parducci, Marshall, & Degner, 1966), lengths of lines (Restle & Levison, 1971), loudness of tones (Marks, 1992; Parducci & Sandusky, 1970; Schneider & Parker, 1990), and even psychopathy of verbal materials (Manis, Nelson, & Shedler, 1988).

Two types of theoretical accounts of these effects have been described in the literature. Here I refer to the first of these as a stimulus-based account, because this account implies that subjective values of the stimuli being compared are altered by an implicit process that compares each stimulus with features of its respective contextual distribution. According to this framework, shifts in choice proportions reflect shifts in perceptions of the magnitudes of the different stimuli. Stimulus-based accounts of contrast effects in paired comparisons have been framed in terms of Helson's (1964) adaptation-level (AL) theory (Helson et al., 1954; Marks & Warner, 1991; Restle & Levison, 1971), which assumes that the stimulus is valued in terms of its deviation from the mean of contextual stimuli.

The second type of theoretical account ties contrast effects in paired comparisons to response tendencies rather than to the features of the stimulus distribution directly. For example, Erlebacher and Sekuler (1971) argued that when

[1] Paired comparisons involve a dominance-based comparative judgment in which the participant chooses which of two stimuli has a greater value on the specified attribute dimension. Thus, paired comparisons may be viewed as both a choice and a judgment task. I use these terms interchangeably throughout the article.

participants are uncertain about which stimulus is of greater magnitude, they will tend to respond in a way that will equalize the frequency of using each response category. On the basis of this analysis, Erlebacher and Sekuler argued that context-induced changes in PSE were an artifact of the method of paired comparisons rather than a result of changes in perception.

Although both stimulus-based and response-based models have provided plausible explanations of contrast effects in paired comparisons, there have been no definitive tests of these models. The purpose of the experiments reported here was to provide such tests. These experiments tested both qualitative and quantitative predictions of three specific models. Two of these models are stimulus based, and one is response based. The first of the stimulus-based models is derived from AL theory, which explains contrast effects in paired comparisons as resulting from differences in the means of relevant stimulus distributions. An alternative stimulus-based model is derived from range-frequency theory (Parducci, 1983), which conceives of context effects as reflecting differences in stimulus ranks and differences in stimulus values relative to the highest and lowest valued stimuli in the contextual distribution. I refer to this instantiation of range-frequency theory as the rank-dependent valuation model because it predicts differential context effects for different stimulus pairs depending on the ranks of the stimuli in their respective contexts.

The third model can also be viewed as being derived from range-frequency theory. Although Parducci (1983) has argued that range-frequency theory may operate on subjective valuations, he has also presented a version of the theory in which context effects are derived from response tendencies (Parducci, 1965). For current purposes, the critical response tendency described by the model is the tendency to use categories with equal frequency (the frequency principle). I refer to this model as the response-equalization model.

In the next several sections, I describe these models in detail so that quantitative fits to the data can be generated from them. I then describe three experiments that tested different predictions derived by the models.

Logistic Representation of the Choice Process

There is a long history of modeling choice proportions generated from paired-comparisons tasks as a monotonic function of the differences in the scale values of the stimuli being compared. Different researchers have proposed different monotone functions, such as the cumulative normal ogive (Thurstone, 1927a) and the logistic ogive (Luce, 1959). Link (1992) has made a strong case for using the logistic function on the basis of a theory in which the stimulus signal is represented by a continuous wave form that is sampled across time. The logistic distribution falls out of an assumption that the waveform is generated by a Poisson process. Furthermore, the theory directly ties choice proportions to response latencies through modeling the choice process in terms of a random walk. For these reasons, I use the logistic function to model the choice process.

The use of a common function for relating model parameters to choice proportions also permits a more direct comparison of the predictive fit of the three models.

According to the logistic model, the probability that a participant will choose stimulus X_{i1} over X_{j2} is given by

$$p(X_{i1} > X_{j2}) = 1/(1 + e^{-b(J_{i1}-J_{j2})}), \quad (1)$$

where J_{i1} and J_{j2} represent the internal values of the two stimuli in their respective contexts and the multiplicative constant b represents a discriminability parameter that affects the slope of the probability function. In the context-independent model, values for internal judgments do not depend on the context from which the stimulus was drawn, and so the choice proportions for cases in which a stimulus is compared with itself should equal .5.

AL Model

The basic premise behind AL theory (Helson, 1964) is the conception that the perceptual system is geared to adapt to changing conditions and that perceived value on an attribute dimension is determined relative to a neutral point representing the organism's level of adaptation. For example, when one has been out in cold weather (temperatures below freezing), tepid water may be perceived as warm or hot to the touch. On the other hand, if one has been out in very hot weather, that same water may be perceived as cool or cold. Presumably, exposure to the cold weather lowered the AL for temperature so that the water temperature was perceived as warmer than neutral, but exposure to hot weather raised the AL so that the water temperature was perceived as cooler than neutral.

Helson (1964) argued that the type of adaptation mechanism just described could be used to explain contrast effects typically observed in absolute judgment. The basic contrast effect in judgment corresponds to a tendency to rate a stimulus value high on the judgment scale when the experimental set of stimuli consists mostly of lower valued stimuli and to rate it low when the set consists of mostly higher valued stimuli. To account for this effect, the AL must depend on the experimental series of stimuli. In particular, Helson (1947) provided evidence that the empirical AL, which was defined as the stimulus value rated neutral on the scale, typically corresponded to the mean (or sometimes geometric mean) of stimulus values. Thus, the AL can be conceived as a moving average that takes into account the recently experienced context. However, Helson reasoned that the AL was not strictly determined by the experimental series. Instead, it could be conceived as a weighted average of three sources of stimulation: (a) the values in the experimental series, (b) background values, and (c) residual values that the participant might carry into the experiment. Equation 2 formalizes this idea by representing the AL in context k as a weighted average of the mean of context-independent stimulus scale values (S), background values (B), and residual values (R):

$$AL_k = (z_1\bar{S}_k + z_2\bar{B}_k + z_3\bar{R}_k)/(z_1 + z_2 + z_3). \quad (2)$$

The AL may then be used to generate an internal judgment of the stimulus. In Equation 3, the subjective judgment (J_{ik}) of stimulus i in context k on a scale bounded by 0 and 1 is conceived as a linear function of the difference between the context-independent value of that stimulus (S_i) and the AL:

$$J_{ik} = .5 + m(S_i - AL_k) . \qquad (3)$$

The additive constant is set to .5 because the stimulus should be judged neutral on the scale (.5) whenever its scale value equals the AL.

When two stimuli are compared, the difference in their subjective magnitude is then a function of the difference in context-free scale values and ALs:

$$J_{i1} - J_{j2} = m[(S_i - S_j) + (AL_2 - AL_1)] . \qquad (4)$$

If one assumes that background and residual factors are the same for the two contextual sets, then the difference judgments can be described entirely in terms of the context-independent scale values:

$$J_{i1} - J_{j2} = m[(S_i - S_j) + z(\bar{S}_2 - \bar{S}_1)] . \qquad (5)$$

As formulated in Equation 5, AL theory provides predictions of difference judgments on the basis of stimulus information alone. Once the psychophysical scale relating physical properties of the stimulus to scale values is established, the model reduces to a simple weighted additive combination of two sources: the differences between the scale values being compared and the difference between the ALs of the distributions from which they were selected.

To estimate choice proportions, one substitutes the parameters of Equation 5 into the logistic function as follows:

$$p(X_{i1} > X_{j2}) = 1/(1 + e^{-b[S_i - S_j + z(\bar{S}_2 - \bar{S}_1)]}) . \qquad (6)$$

The parameter m disappears because it is absorbed in the slope parameter b. If the psychophysical function relating physical stimulus values (X) to scale values (S) is known, then only two parameters (b and z) must be fitted to the data. When $z = 0$, the model reduces to the context-independent model. The predictive power of the model is enhanced to the degree to which the fitted parameters are constant across manipulations of context.

Rank-Dependent Valuation Model

In the domain of absolute judgment, AL theory has been proven to be inferior to range-frequency theory in predicting the effects of manipulating context (Birnbaum, 1974; Parducci, 1963; Parducci, Calfee, Marshall, & Davidson, 1960; Parducci & Perrett, 1971). This conclusion is based on results from experiments in which AL was held constant across contexts but context effects were still obtained. Range-frequency theory has successfully modeled these contextual effects by assuming that the form of the rating function depends on the form of the cumulative frequency function for the contextual distribution. Thus, a distinguishing feature of range-frequency theory is that it ties contextual effects to rank information. Although range-frequency theory has been demonstrated to be superior to AL theory in predicting absolute judgments, this does not mean that one can assume that this superiority will carry over into the realm of paired comparisons. For example, reference points and aspiration levels have played an important role in theories of choice (Kahneman & Tversky, 1979; Seigel, 1957). Thus, it is possible that the AL may take on a greater role in a choice situation than in the absolute judgment situation.

The rank-dependent valuation model is an instantiation of range-frequency theory in which rank-order effects operate at the valuation stage rather than at a response selection stage in the judgment process. The effects of manipulating ranks are captured by the frequency principle, according to which the value of a stimulus depends on its percentile rank in the relevant contextual distribution. The frequency value of stimulus i in context k is given as

$$F_{ik} = (\text{rank}_{ik} - 1)/(N_k - 1) , \qquad (7)$$

where N_k is the number of stimuli in the contextual distribution. The internal judgment of a stimulus is assumed to be a weighted average of range and frequency values, with the range value of a stimulus equal to the proportion of the subjective contextual range lying below that stimulus. Thus, the internal judgment of stimulus i in context k is given as

$$J_{ik} = w[(S_i - S_{\min, k})/(S_{\max, k} - S_{\min, k})] + (1 - w)F_{ik} , \qquad (8)$$

where $S_{\max,k}$ and $S_{\min,k}$ are the maximum and minimum values defining the subjective range and w is the weighting constant.

If one assumes that a common range is used to evaluate stimuli in pairwise judgments, then the difference in internal judgments is represented as

$$J_{i1} - J_{j2} = w[(S_i - S_j)/\text{range}] + (1 - w)(F_{i1} - F_{j2}) , \qquad (9)$$

where the range is equal to the difference between the highest value of $S_{\max, k}$ and the lowest value of $S_{\min, k}$ across contexts. In the rank-dependent valuation model of Equation 9, the contrast effects in paired comparisons are due to differences in ranks of stimuli in the two contexts. When $w = 1$, the model reduces to the context-independent model.

To estimate choice proportions, one substitutes the parameters of Equation 9 into the logistic function as follows:

$$p(X_{i1} > X_{j2}) = 1/(1 + e^{-b(w[(S_i - S_j)/\text{range}] + (1 - w)(F_{i1} - F_{j2}))}) . \qquad (10)$$

The rank-dependent valuation model is similar to the AL model in that the locus of the context effect is at the subjective valuation of each stimulus. The critical difference is that the AL model predicts constant effects of context, whereas the rank-dependent valuation model predicts that the effects of context will vary across stimuli if rank differences likewise vary. Wedell (in press) has provided some support for the rank-dependent model in pairwise similarity ratings of squares and dot patterns. However, the processes involved in similarity judgments may

differ in critical ways from those involved in paired comparisons.

Equation 10 has three parameters that must be fit to the data (i.e., b, w, and range) as compared with the two parameters of the AL model described earlier and the response-equalization model described subsequently. As a means of facilitating comparisons with these other two models, the following simplified version of the rank-dependent model is used in fitting the data:

$$p(X_{i1} > X_{j2}) = 1/(1 + e^{-b[S_i - S_j + z(F_{i1} - F_{j2})]}). \quad (11)$$

Using Equation 11 to represent the rank-dependent valuation model emphasizes the placement of context effects at the level of rank differences and has the advantage of equating the number of fitted parameters across the three models.

Response-Equalization Model

An alternative way to describe the frequency principle is as a tendency for participants to use response categories with equal frequency (Parducci, 1965). This characterization of the frequency principle is consistent with an interpretation of context effects as resulting from the response selection process rather than the valuation process. Under this interpretation, the subjective difference in the values of two stimuli does not depend on the context, but contrast effects may still be observed because of a shift in the criterion used to define the different comparative judgment categories. In particular, the criterion location will depend on the relative frequency with which each response category has been assigned to stimulus comparisons across the two contexts.

To generate predictions of these effects from the stimulus distributions, I introduce a statistic, RP_{km} (RP refers to response proportion), that represents the proportion of stimulus comparisons strictly favoring context k over context m. For example, let Stimulus Context 1 be defined as the set of values $\{1, 2, 3, 4\}$ and Stimulus Context 2 be defined as the set of values $\{3, 4, 5, 6\}$. Assuming that all 16 pairwise combinations of the two sets are constructed, the value of RP_{12} is calculated as 1/16 because only 1 of the 16 comparisons strictly favors a stimulus from Context 1 over a stimulus from Context 2 (i.e., $4 > 3$). The value of RP_{21} is 13/16, which is calculated by noting that Stimulus 3 is greater than two of the stimuli in Context 1, Stimulus 4 is greater than three of the stimuli in Context 1, and Stimuli 5 and 6 are each greater than all four stimuli from Context 1. Thus, the RP values reflect the proportion of comparisons strictly favoring stimuli from one context over stimuli from the other context.

Equation 12 can then be used to estimate choice proportions for the response-equalization model as follows:

$$p(X_{i1} > X_{j2}) = 1/(1 + e^{-b[S_i - S_j + z(RP_{21} - RP_{12})]}). \quad (12)$$

When RPs for the two contexts are equal, no context effects should occur. Similarly, no context effects should occur if the weighting of the differences in response proportions is

zero. The model presented in Equation 12 differs from that presented by Erlebacher and Sekuler (1971) in two ways. First, the model assumes a logistic transformation function rather than a normal ogive. Second, it uses a single response cutoff, whereas their model assumed two response cutoffs, with the response-equalization principle operating only in the region of response uncertainty.

Overview of Experiments

The experiments reported here were designed as tests among the three theoretical models described in Equations 6, 11, and 12. Each of these models has the same form, describing paired-comparisons probabilities as a logistic function of differences in context-independent scale values and differences in contextual parameters. They differ in the nature of the contextual parameters. Both the AL model and the rank-dependent valuation model tie the contextual parameters to features of the stimulus distribution, whereas the response-equalization model ties the contextual parameters to responses associated with the paired distributions. The rank-dependent model stands out as the only one of these models in which context effects can vary for pairs of stimuli in which the difference in scale values is the same.

Experiment 1 manipulated the difference in ALs while holding both rank differences and differences in RPs constant. Sensitivity to this manipulation would provide some support for the AL model over the other two models. However, if the quantitative fits to the data indicated the need for changes in the weighting of differences in ALs to account for these effects, then the support for the AL model would be considerably weakened because similar changes in the weighting parameters of the other models would provide equivalent fits. The focus of Experiment 2 was on testing the unique prediction of the rank-dependent valuation model that contrast effects can vary across stimulus pairs that have been equated on differences in scale values. Confirmation of this prediction would constitute strong support for the rank-dependent model. In Experiment 3, differences in RPs and ALs were manipulated independently. One pair of distributions was constructed to equate ALs and vary RPs, and a second pair of distributions was constructed to vary ALs and equate RPs. This experiment provided a critical test between these two theoretical accounts.

Experiment 1: Manipulating ALs for Simultaneous and Successive Comparisons of Square Size

Experiment 1 had three major goals. The primary goal was to establish that the predicted contrast effects occurred in paired comparisons of square sizes. Previous research had demonstrated contrast effects on paired comparisons of line sizes (Restle & Levison, 1971), and so generalization to square size comparisons was expected. A second goal was to examine a procedural issue, namely, whether the contrast effects in paired comparisons would occur for both simultaneous and successive presentation methods. The experi-

ments in the literature reviewed earlier all used a successive presentation method in which one member of the pair was presented first and then, after a short delay, the second member of the pair was presented. Experiment 1 used both the successive presentation method and a simultaneous presentation method in which both squares were flashed on the computer screen at the same time. Previous research has demonstrated that context effects on category ratings tend to be larger with simultaneous presentation of stimuli (Parducci & Wedell, 1986; Wedell & Parducci, 1988). However, research on pairwise dissimilarity judgments indicates that context effects may be minimized when both members of a pair are viewed simultaneously (Mellers & Birnbaum, 1982; Wedell, in press).

A third goal was to provide a test of predictions that were unique to the AL model. This was achieved by constructing two pairs of distributions. In the first pair of distributions, one distribution had a lower AL, a lower RP, and higher ranks for common stimuli. The second pair of distributions was constructed such that differences in ranks and RPs were the same as for the first pair of distributions but differences in ALs increased dramatically. It should be emphasized that observation of increased context effects for this second distribution would provide strong support for the AL model only to the extent that the quantitative fits are based on a constant weighting parameter. This is because all three models can account for differences between these two conditions if the weighting parameters are allowed to vary.

Method

Participants. Participants were 78 undergraduates from the University of South Carolina who received course credit for their participation. A 2 × 2 × 2 × 25 factorial design was used. The between-subjects variables were distribution set (narrow–high range [NH] vs. narrow–low range [NL] or wide–high [WH] vs. wide–low [WL]), presentation mode (simultaneous vs. successive), and distribution arrangement (varying which distribution was presented on the left or right of the screen). Students were randomly assigned to one of the eight between-subjects conditions. The 25 levels of the within-subjects variable constituted the 25 different pairings of stimuli. Each student encountered each pair on 8 trials. The presentation order of these 25 trials was randomized within each of eight blocks for each student separately. The dependent variable was the proportion of times the stimulus from the low distribution was judged larger than the stimulus from the high distribution (based on 8 trials per stimulus pair).

Apparatus. All stimuli and instructions were presented on IBM/PS2 50Z computers with 10-in. (25.4-cm) IBM color monitors. Stimuli were presented in VGA graphics mode on a white background, with the width of a square pixel equal to 0.3125 mm.

Stimuli. Stimuli were selected from a series of 25 square sizes varying from 10 to 250 pixels in width, with a 10-pixel difference in the widths of adjacent sizes. Table 1 presents the stimulus distributions used in the three experiments. As shown in Table 1, three square sizes, 9, 10, and 11, were common to each pair of distributions used in Experiment 1. The two additional stimuli making up a distribution were of higher value than the target stimuli in the NH and WH sets and of lower value in the NL and

Table 1
Stimulus Conditions of Experiments 1–3

Experiment/ distribution	Square sizes	Adaptation level	Response proportion	Frequency values for common stimuli		
1						
NL	7, 8, *9,10, 11*	9.00	.120	0.50	0.75	1.00
NH	*9,10,11,*12, 13	11.00	.760	0.00	0.25	0.50
Difference		−2.00	−.640	0.50	0.50	0.50
WL	1, 5, *9,10, 11*	7.20	.120	0.50	0.75	1.00
WH	*9,10,11,*15, 19	12.80	.760	0.00	0.25	0.50
Difference		−5.60	−.640	0.50	0.50	0.50
2						
EB	6, 9,10,11,*12,15*	10.50	.444	0.00	0.20 0.80 1.00	
EU	6, 7, *9,12,*14,*15*	10.50	.444	0.0	0.40 0.60 1.00	
Difference		0.00	.000	0.00	−0.20 0.20 0.00	
EP	6, 7, 8, *9,12,15*	9.50	.278	0.00	0.60 0.80 1.00	
EN	6, *9,12,*13,14,*15*	11.50	.639	0.00	0.20 0.40 1.00	
Difference		−2.00	−.361	0.00	0.40 0.40 0.00	
3						
AL	3, 4,*10,11,*12,19	9.83	.472	0.40	0.60	
AH	2, 9,*10,11,*17,18	11.50	.472	0.40	0.60	
Difference		−1.67	.000	0.00	0.00	
HP	6, 7, 8, *9,11,*20	10.17	.333	0.60	0.80	
LN	1, *9,11,*12,13,15	10.17	.611	0.20	0.40	
Difference		0.00	−.278	0.40	0.40	

Note. Values in italics are common to both the paired distributions. NL = narrow–low range; NH = narrow–high range; WL = wide–low range; WH = wide–high range; EB = equal-range bell distribution; EU = equal-range U-shaped distribution; EP = equal-range positive skew; EN = equal-range negative skew; AL = adaptation level low; AH = adaptation level high; HP = high-range positive skew; LN = low-range negative skew.

WL sets. The stimuli in each distribution covered a range of 5 square sizes in the narrow-range conditions and 11 square sizes in the wide-range conditions. The full set of stimulus pairs was constructed by pairing each stimulus from the first distribution with each stimulus from the second distribution. Stimuli from the first distribution appeared on the left side of the screen in blue, and stimuli from the second distribution appeared on the right side of the screen in red.

Procedure. Up to 6 students at a time were tested in a large room with terminals spaced along the outer wall at approximately 1-m intervals. After filling out informed-consent forms, students read instructions on the computer that told them that their task was to answer as quickly and accurately as possible which of two squares presented during a trial was larger. If they thought that the left square appeared larger, they were to press the *1* on the keypad. If they thought that the right square appeared larger, they were to press the *2* on the keypad. They were instructed to place the index and middle fingers of their right hand on the corresponding numbers of the keypad to facilitate fast and accurate responses. A verbal reminder of which numeral was associated with which response category appeared at the bottom of the screen throughout the experiment.

For students in the simultaneous presentation condition, a trial began with the word *ready* printed in the center of the screen for 1 s as a prompt, followed immediately by the presentation of both squares for 1 s. This was followed by an unfilled interval until a response was registered. Response latencies were measured from the onset of the pair of squares until a response key was pressed. Responses were printed on the screen for 0.5 s. If the response was not legitimate, the student was prompted to provide a legitimate response (however, there were no occurrences of illegitimate responses on any trials in the experiments reported here). There was a 1.5-s unfilled interval before the beginning of the next trial.

For students in the successive presentation condition, the procedure for a trial was nearly the same as in the simultaneous presentation condition. The only exceptions were that the blue square from the first distribution was presented first for 1 s on the left side of the screen, followed by a 2-s unfilled interval and, finally, a 1-s presentation of the red square from the second distribution on the right side of the screen. Response latencies were measured from the onset of the second square until a response key was pressed.

Results

The effects of the contextual manipulation can be gauged by examining choice proportions for stimuli common to each pair of distributions. There were three common stimuli (Squares 9, 10, and 11). From these, one can examine zero-step comparisons of a stimulus with itself (e.g., 9 vs. 9), one-step comparisons (e.g., 9 vs. 10), or two-step comparisons (e.g., 9 vs. 11). Table 2 presents these choice proportions for each between-subjects experimental condition. Choice proportions are coded to reflect the proportion of times the contextually favored target (e.g., the square from the low distribution) was judged larger.

Analysis of choice proportions. Consider first the zero-step comparisons. According to the null hypothesis of contextual independence, the expectation for these choice proportions is .5. Instead, the results reflect the predicted contrast effects on choice, with choice proportions strongly favoring the square drawn from the low-range distribution. A one-tailed z test of a single proportion was conducted on each of the 24 zero-step comparisons (at the $p = .05$ level). All tests were significant except for the 9-9 comparison in the WL-WH, 2-s delay condition and the 11-11 comparison in the NL-NH, 2-s delay condition. Considering that the

Table 2

Proportion of Judgments Indicating the Target Square From the Low Distribution Higher: Experiment 1

Stimulus pair[a]	0-s delay				2-s delay			
	Narrow range		Wide range		Narrow range		Wide range	
	Low–high ($n = 10$)	High–low ($n = 11$)	Low–high ($n = 9$)	High–low ($n = 10$)	Low–high ($n = 10$)	High–low ($n = 9$)	Low–high ($n = 9$)	High–low ($n = 10$)
Zero step								
9-9	.913	.864	.958	.913	.663	.806	.611	.725
10-10	.875	.932	.958	.963	.675	.875	.653	.713
11-11	.888	.909	1.000	.975	.650	.819	.667	.788
M	.892	.902	.972	.950	.663	.833	.644	.742
One step down								
10-9	1.000	1.000	1.000	1.000	.900	.944	.917	.863
11-10	1.000	.977	.972	.988	.888	.986	.958	.900
M	1.000	.989	.986	.994	.894	.965	.938	.882
One step up								
9-10	.100	.227	.222	.338	.300	.361	.250	.350
10-11	.188	.273	.194	.225	.225	.451	.250	.400
M	.144	.250	.208	.282	.263	.410	.250	.375
Two steps down								
9-11	1.000	.989	.986	1.000	.975	.992	.958	.938
Two steps up								
11-9	.000	.023	.098	.038	.038	.069	.083	.088

[a] The first number is the square size from the low distribution, and the second number is the square size from the high distribution.

average standard error of a proportion was .035, the observed context effects on zero-step comparisons were quite large.

The effects of context on zero-step comparisons were smaller for the successive presentation condition. Within the successive presentation conditions, they were most strongly reduced when the low-range distribution was presented first. As a means of statistically testing these differences across conditions, a $2 \times 2 \times 2 \times 3$ repeated measures multivariate analysis of variance (MANOVA) was conducted on the 24 conditions shown for the zero-step comparisons of Table 2. The repeated measures variable consisted of the three stimulus pairs, and the between-subjects variables were distribution set, presentation mode, and distribution arrangement. The analyses reported subsequently used the mean proportion for each student as the dependent variable. The same analyses were conducted on the arcsine transformation of those proportions. Any differences in the pattern of significance from these two analyses are reported.

There was no significant main effect of stimulus pair, and there were no significant interactions of stimulus pair with the between-subjects variables ($ps > .25$). Thus, there was no evidence of differential effects of context across stimulus pairs. The main effect of presentation mode was significant, $F(1, 70) = 42.40, p < .001, MSE = 0.060$, reflecting much greater effects of context for the simultaneous presentation condition. There was also a significant two-way interaction of presentation mode with distribution arrangement, $F(1, 70) = 4.82, p < .05, MSE = 0.060$, and a marginally significant two-way interaction of presentation mode with distribution set, $F(1, 70) = 3.50, p < .07, MSE = 0.060$. These interactions indicated the need for examining the effects of context separately for simultaneous and successive presentations.

For the simultaneous presentation condition, the only significant finding was the main effect of distribution set, $F(1, 36) = 4.12, p = .05, MSE = 0.030$. This result reflected the significantly greater effects of context for the wide-range distribution conditions in which the difference in ALs was greater. For the successive presentation condition, the only significant effect was for distribution arrangement, $F(1, 34) = 5.64, p < .05, MSE = 0.091$, reflecting significantly greater effects of context when the squares from the high distributions were presented first.[2] This effect can be explained as the result of a constant negative time order error in which the remembered size of the first stimulus presented decreases with time. A similar negative time order error was postulated by Erlebacher and Sekuler (1971) in modeling context effects on paired comparisons of line length. More generally, negative time order errors have typically been observed in paired comparisons for psychophysical stimuli (Helson, 1964).

In addition to analyzing zero-step comparisons, one can examine the effects of context on comparisons of squares differing by one or two sizes. Under the null hypothesis of context-independent choice, the choice proportions for one-step-up comparisons should be complementary to the corresponding one-step-down comparisons. In other words, the choice proportions for upward and downward comparisons

shown in Table 2 should add to 1.0. Contrast effects imply that the choice proportions for the contextually favored stimulus would be biased upward in both one-step-down and one-step-up comparisons. Thus, contrast effects are represented by superadditivity of these choice proportions. The choice proportions of Table 2 clearly show such superadditivity for one-step comparisons but not for two-step comparisons. Separate one-tailed t tests of these combined proportions (at $p < .05$, $H_0:\mu = 1.0$) confirmed that all one-step comparisons showed significant superadditivity except for the 9-10 pair in the NH-NL, 0-s delay condition and the 10-11 pair in the NH-NL, 2-s delay condition. None of the two-step comparisons showed significant superadditivity. The lack of context effects at the two-step comparison level is probably a result of the high discriminability between squares differing by two sizes.

A $2 \times 2 \times 2 \times 2$ repeated measures MANOVA with the same between-subjects variables described earlier was run on the summed proportions of one-step-up and one-step-down data. The only significant result was a main effect of distribution arrangement, $F(1, 70) = 4.56, p < .036, MSE = 0.114$. This effect reflected greater superadditivity when the high-range distribution was presented first on the left side of the screen.

Model fits. The analyses of choice proportions described earlier led to three major conclusions: (a) Context strongly affected choice proportions, (b) these effects were greater for simultaneous presentation, and (c) increasing differences in ALs led to increased context effects in the simultaneous presentation condition but not in the successive presentation condition. This last conclusion suggests that the AL model may have a predictive advantage over the other models, at least in the simultaneous presentation condition. However, all three models can account for this difference if the contextual weighting parameter is allowed to vary as a function of wide-range versus narrow-range conditions. Thus, unique support for the AL model depends on the model being able to account for this effect without requiring a corresponding change in the weighting parameter.

The purpose of the model fits described in this section was to answer the specific question of whether equally constrained models provide comparable fits to the data. To answer this question, I substituted square size (e.g., physical width) for the scale values (S_i) in Equations 6, 11, and 12 on the basis of previous research on square judgment in which this assumption proved plausible (Parducci & Wedell, 1986;

---

[2] The MANOVAs conducted on the arcsine-transformed proportions differed in the following ways. The two-way interaction between distribution set and presentation mode was significant ($p < .05$), but the two-way interaction between presentation mode and presentation order was not ($p > .10$). Simple effects analysis of the simultaneous presentation conditions again revealed a significant effect of distribution set ($p < .05$). Simple effects analysis of the successive presentation condition revealed only a marginal effect of presentation order ($p = .053$). In general, this pattern of results supports the same conclusions based on the analysis of mean proportions described in the text.

Wedell, in press). The equations were fit to the data of the zero-, one-, and two-step comparisons for the eight between-subjects conditions of Experiment 1. Data points corresponding to proportions less than .02 or greater than .98 were first eliminated. This left 13 data points in the simultaneous condition and 36 in the successive presentation condition. The models were fit with a nonlinear estimation procedure. Results using least squares and maximum-likelihood loss functions yielded generally comparable fits.[3]

A convenient index of model fit is the R^2 value, or the proportion of variance in mean proportions explained by the models. The R^2 values reported for the different models are based on the results from the least squares fitting procedure. To evaluate the fit of the models statistically, I subtracted predicted proportions from the observed mean proportions of each student for zero-step and one-step comparisons. I then used this residual proportion as the dependent variable in a $2 \times 2 \times 2 \times 7$ repeated measures MANOVA (I excluded two-step proportions because of the lack of variance in these conditions). In calculating residuals, I used predicted values derived from the maximum-likelihood procedure.

All three models have the same form, with a discriminability parameter (b) and a contextual weighting parameter (z). Because of the observed order effects in successive presentation conditions, a time order error parameter (t) was included as an additive constant in the exponent of each model. When t was 0.0, the predicted choice proportions did not depend on which distribution was presented first. However, the results indicated a negative time order error in the successive presentation conditions. Negative values of t resulted in increasing the choice proportions for the square presented second over the square presented first. Psychologically, this corresponded to the memory for the value of the first square being decremented by a constant.

The three parameters of each model can be fit to data from each of the eight between-subjects conditions (a total of 24 [3×8] possible fitted parameters). The approach to model testing taken here was to begin with the most constrained models and assess whether freeing parameters to fit different conditions significantly increased the model's fit.

Because of the obvious effects of presentation mode on choice proportions, the simplest context-independent model that was fit had three free parameters. This model, in which the discriminability parameter (b) varied with presentation mode and a nonzero time order constant (t) was fit to the successive presentation conditions, left a substantial proportion of the variance in mean proportions unexplained $(R^2 = .62)$. As a means of improving on this fit, the simplest four-parameter context-dependent models included a single value of the context weighting parameter, z. The four-parameter AL model greatly increased the proportion of variance explained over the context-independent model $(R^2 = .92)$. However, the four-parameter response-equalization model provided an even better fit to the mean proportions $(R^2 = .98)$. Despite the better fit, the four-parameter response-equalization model could not account for the Distribution Set $\times$ Presentation Mode inter-

action found in the data because it predicted equal effects of contexts for both the wide-range and narrow-range conditions. A parsimonious explanation of this effect within the AL model would be to allow z to vary with presentation mode; however, this five-parameter model did not substantially improve the fit of the model $(R^2 = .92)$. The problem with the model is that it overpredicted the effects of manipulating range in the simultaneous presentation condition. A much better five-parameter fit is achieved with the AL model by allowing z to vary with stimulus set $(R^2 = .98)$. However, this model is exactly the same as the corresponding five-parameter response-equalization model, with z varying with distribution set. The reason is that the critical difference between the AL model and the other models was that AL model was the only one that predicted differences in proportions resulting from distribution set (when z was held constant for the different sets). Allowing z to vary with distribution set means that, for all models, the product of the fitted z and the difference in contextual parameters (ALs, RPs, or Fs) will be the same.

The upshot of these analyses is that although the AL model had a potential advantage over the other two models in predicting greater effects of context for the wide-range sets in the simultaneous presentation condition, it did a poor job unless the weighting of the AL difference was allowed to vary with range. However, when z varies with range, the three models are equivalent and so cannot be distinguished. One might argue that the poor fit of the AL model was due to the use of arithmetic means rather than geometric means (as originally formulated by Helson, 1964) to represent ALs. However, it should be noted that the difference in ALs increased rather than decreased when computed on the basis of geometric means. Thus, the four-parameter AL model based on geometric means provided an even poorer fit to the data than the equivalent model based on arithmetic means.

Finally, one may wish to assess how well the models fit the data using the more sensitive repeated measures MANOVA on residuals. Models based on up to nine parameters all yielded significant effects, and so none adequately explained the full pattern of data. However, it is instructive to examine how the models misfit. To this end, the fit of the nine-parameter response-equalization model is presented in Table 3. The nine fitted parameters are one value of t for successive presentations and separate values of b and z for each of the four Presentation Mode $\times$ Contextual Set conditions. This model is equivalent to the nine-parameter AL model, but the contextual parameters are more easily interpreted. The model resulted in only two

---

[3] The maximum-likelihood loss function was:

Loss $= $ count $\cdot$ {$p \cdot$ log(estimate) $+ [1 - p \cdot$ log(1 $-$ estimate)]},

where count is the number of observations making up each data point. Because each Presentation Mode $\times$ Distribution Set condition consisted of approximately 20 students who each contributed eight choices to each data point, the count was typically 160 (20 $\times$ 8). In general, the estimates of the discriminability parameter, b, were slightly smaller for the maximum-likelihood method than for the least squares method.

Table 3

Fit of the Nine-Parameter Response-Equalization Model to the Mean Zero-Step and One-Step Proportions: Experiment 1

	Parameter			Residual[a]						
Condition	t[b]	b	z	9-9	10-10	11-11	9-10	10-11	10-9	11-10
NL-NH, 0 s	0.000	3.433	0.814	.081	.043	.056	.008	.008	−.063	.024
NH-NL, 0 s	0.000	3.433	0.814	−.034	.035	.012	.005	−.018	−.029	.016
WL-WH, 0 s	0.000	3.576	0.997	.052	.052	.093	.003	−.025	.008	−.020
WH-WL, 0 s	0.000	3.576	0.997	−.036	.014	.027	.002	−.011	−.001	−.114
NL-NH, 2 s	−0.173	1.713	1.038	−.030	−.017	−.042	−.003	−.015	−.053	−.128
NH-NL, 2 s	−0.173	1.713	1.038	.063	.132	.077	.022	.064	−.051	.046
WL-WH, 2 s	−0.173	1.493	0.915	−.030	.011	.025	.093	.135	.093	−.156
WH-WL, 2 s	−0.173	1.493	0.915	.046	.033	.108	.015	.053	−.098	−.048

Note. NL = Narrow–low range, NH = narrow–high range; WL = wide–low range; WH = wide–high range.
[a] The first number is the square size from the low distribution, and the second number is the square size from the high distribution. [b] The t parameter was fixed at zero for simultaneous presentation conditions.

significant effects, a main effect of presentation order and an interaction of presentation order with stimulus pair.

Examination of the residuals indicates that the model fit the simultaneous presentation data quite well, with only 3 of 28 predicted proportions deviating by more than .07 (the average standard deviation was approximately .035). The misfit of the model appeared to be localized in the one-step comparisons for the wide-range, successive presentation conditions, for which 5 of 8 predictions deviated by more than .09.

The fitted parameter values provide some insights into the pattern of results. First, the contextual weighting increased with the range manipulation in the simultaneous presentation condition but actually decreased slightly in the successive presentation condition. Thus, the effects of manipulating AL were somewhat different under the two presentation modes. Second, the large main effect of presentation mode on the magnitude of contextual effects was not modeled in terms of greater values of z for simultaneous presentation conditions; rather, it was due to the greater discriminability in the simultaneous presentation condition (i.e., greater values of b). Within the logistic model, as discriminability decreases, proportions regress toward .5, reducing the magnitude of context effects even when the contextual weighting is held constant. The decreased context effects in the successive presentation condition appear inconsistent with the response-equalization model of Erlebacher and Sekuler (1971), which assumes that the response-equalization tendency occurs only within a region of uncertainty. Presumably, the poorer discriminability associated with successive presentations should increase the region of uncertainty and lead to greater, not lesser, effects of context.

Discussion

Experiment 1 demonstrated strong contextual effects on paired comparisons of square sizes. These effects were large for both simultaneous and successive presentation conditions. Overall, the effects of context were greater for simultaneous presentation, but these effects were modeled as the result of increased discriminability rather than increased

weighting of contextual information. In addition to these effects, the pattern of results differed for the two presentation modes. Increasing the differences in stimulus means of the distributions being compared resulted in a corresponding increase in context effects only for the simultaneous presentation conditions. Although the AL model predicts this type of increase, model fits required that the contextual weighting parameter also vary with range. Under these conditions, both the rank-dependent valuation and response-equalization models provide equivalent fits. Thus, the three models predicted judgment equally well in Experiment 1.

Experiment 2: Frequency Manipulations Equating Ranges

Experiment 2 was designed as a strong test between the rank-dependent valuation model and the other two models. Because the AL model and the response-equalization model tie the context-dependent parameters to a single feature of the stimulus distributions (difference in stimulus means or difference in response proportions), they predict constant effects of context given the same difference in stimulus values. Because the rank-dependent valuation model attributes the effects of context to differences in stimulus ranks, context effects can vary for the same difference in stimulus values. For example, it is possible to equate ranks for some stimulus values and not others across the two distributions being compared. The rank-dependent valuation model predicts no context effects for comparisons of the same stimuli when ranks are equated but strong context effects for comparisons of the same stimuli with different ranks.

Experiment 2 examined these predictions for two pairs of distribution. For the comparisons in which bell-shaped and U-shaped distributions were paired, the AL model and the response-equalization model predicted no effects of context because ALs and RPs were equated for these distributions. The rank-dependent model predicted no effects of context for two of the common stimuli and opposing effects of context for the other two common stimuli. For the compar-

isons in which positively and negatively skewed distributions were paired, the AL model and the response-equalization model predicted large and constant effects of context for the four common stimuli. The rank-dependent model predicted large effects of context for two of the common stimuli and no effects of context for the other two common stimuli. Demonstration of the distinctive patterns of effects described by the rank-dependent model would constitute strong evidence for that model.

Method

The major differences between the methods of Experiments 1 and 2 were in the stimulus sets that were paired together. The apparatus and instructions were the same for the two experiments. The procedure was the same as that used for the simultaneous presentation conditions of Experiment 1.

Participants. Participants were 105 undergraduates from the University of South Carolina who received course credit for their participation. A $2 \times 2 \times 36$ factorial design was used. The between-subjects variables were distribution set (bell shaped vs. U-shaped or positively skewed vs. negatively skewed) and distribution arrangement (bell or positive distributions on the left of the screen or on the right of the screen). Students were randomly assigned to one of the four between-subjects conditions. The 36 levels of the within-subjects variable constituted the 36 different pairings of stimuli. Each student encountered each pair on six trials. The computer randomized the presentation order of these 36 trials within each of six blocks for each student separately. The dependent variable was the proportion of times the stimulus from the U-shaped or positively skewed distribution was judged larger (on the basis of six trials per stimulus pair).

Stimuli. As shown in Table 1, four square sizes (6, 9, 12, and 15) were common to each pair of distributions used in Experiment 2. The two additional stimuli making up each distribution were at Ranks 2 and 5 in the U-shaped distribution, 3 and 4 in the bell-shaped distribution, 2 and 3 in the positively skewed distribution, and 4 and 5 in the negatively skewed distribution. The full set of stimulus pairs was constructed by pairing each stimulus from the first distribution with each stimulus from the second distribution. Stimuli from the first distribution appeared on the left side of the screen in blue, and stimuli from the second distribution appeared on the right side of the screen in red.

Results

Comparison of bell- and U-shaped distributions. Table 4 presents the proportion of times the square from the U-shaped distribution was judged larger than the same square from the bell-shaped distribution. The rank-dependent valuation model predicted

proportions of .5 for the extreme stimulus pairs, in which there should be no effects of context; a proportion greater than .5 for the 9-9 comparison, which favors the U-shaped distribution; and a proportion less than .5 for the 12-12 pair, which favors the bell-shaped distribution. The AL and response-equalization models predicted proportions of .5 for all pairs, indicating no effects of context. The mean proportions averaged across arrangement conditions were close to .5, favoring the latter two models. Two-tailed z tests (at $p = .05$) conducted to test whether any of the four mean proportions differed from .5 were all nonsignificant.

A 2×4 repeated measures MANOVA was conducted on the mean choice proportions for the four pairs of stimuli, with distribution arrangement as the between-subjects variable. Both main effects were significant ($p < .05$) but must be interpreted in light of a significant interaction, $F(3, 49) = 5.23$, $p < .01$. Post hoc trend analyses (using Scheffé's correction) indicated that only the linear trend of target pair interacted with distribution arrangement, $F(1, 51) = 14.08$, $p < .05$, $MSE = 0.055$. This interaction reflects the tendency to judge the square from the U-shaped distribution larger when the bell-shaped distribution was presented first for Squares 6 and 9 but not for Squares 12 and 15. The main effect of stimulus pair was not of the form predicted by the rank-dependent model. The restriction of this effect to the arrangement in which the bell distribution was presented on the left side of the screen suggests that it may be of limited importance.

The comparison most critical to the test of the rank-dependent model is the test of the cubic component. This test was not significant, $F(1, 51) = 0.18$, $p > .50$, $MSE = 0.03$. Failure to reject the null hypothesis brings up the question of the power of the test. Assuming the parameter values fit for the rank-dependent model to data of the wide-range, simultaneous presentation conditions of Experiment 1, the predicted proportions were .50, .60, .40, and .50, respectively, for the four one-step comparisons, 6-6, 9-9, 12-12, and 15-15. The power of the test of the cubic trend, on the basis of these predicted proportions and the obtained mean square error, was greater than .99. Given the high power of the statistical test, the failure to find a significant difference implies a problem with the rank-dependent valuation model.

Comparison of positively and negatively skewed distributions. Table 5 presents the proportion of times the square from the positively skewed distribution was judged larger than the same square from the negatively skewed distribution. The rank-dependent valuation model predicted proportions of .5 for the extreme stimulus pairs, in which there should be no effects of context, and proportions greater than .5 for the middle two pairs, which favor the positively skewed distribution. The AL level and response-equalization models predicted proportions greater than .5 and of similar magnitude for all four pairs, indicating a constant effect of context. The fact that the mean proportions averaged across arrangement conditions were all well above .5 favors the latter two models. One-tailed z tests (at $p = .05$) conducted to test whether each of the four mean proportions differed from .5 were all significant.

A 2×4 repeated measures MANOVA was conducted on the mean choice proportions for the four pairs of stimuli. The only significant effect was the main effect of stimulus pair, $F(3, 48) = 4.29$, $p < .01$. The rank-dependent valuation model predicted a main effect that would be localized in the quadratic trend. Consistent with this prediction, the only significant trend for stimulus pair was the quadratic trend, $F(1, 50) = 9.79$, $p < .01$, $MSE = 0.030$.[4]

---

[4] Repeated measures MANOVAs based on the arcsine transformation yielded the same pattern of significance.

Table 4

Proportions of Judgments Indicating the Target Square From the U-Shaped Distribution Higher: Experiment 2

Stimulus pair	Distribution pairing		
	Bell–U shaped	U-shaped–bell	M
6-6	.698	.436	.567
9-9	.599	.391	.495
12-12	.457	.423	.440
15-15	.426	.468	.447

Table 5

Proportions of Judgment Indicating the Target Square From the Positively Skewed Distribution Higher: Experiment 2

Stimulus pair	Distribution pairing		M
	Positive–negative	Negative–positive	
6-6	.769	.686	.728
9-9	.865	.801	.833
12-12	.859	.750	.804
15-15	.750	.769	.760

Discussion

The results from Experiment 2 provided little support for the rank-dependent valuation model. The model's prediction of the specific pattern of how proportions for same-stimulus comparisons should differ significantly from .5 was not supported. The MANOVA for the bell–*U* sets indicated that the effects of context varied with stimulus pair (at least in one arrangement condition), but these effects were not of the pattern predicted by the rank-dependent model. The MANOVA for the positive–negative sets indicated that the effects of context were also variable, and this time the pattern was consistent with predictions from the rank-dependent valuation model. One interpretation of these effects is that if rank-dependent valuation occurs, these effects are added to other context-dependent processes. Furthermore, such combined effects of rank-dependent processes and more global processes would appear to occur under some distributional conditions but not others. Overall, the evidence for rank-dependent valuation effects on paired comparisons appears weak at best. Although Wedell (in press) has demonstrated the dependence of pairwise similarity judgments on stimulus ranks, manipulation of stimulus ranks does not seem to be the critical determinant of context effects on paired comparisons in the conditions tested in Experiment 2.

Experiment 3: Manipulating ALs and RPs Independently

The purpose of Experiment 3 was to test between the AL and response-equalization models. The most direct test between these models entails manipulating the distributions so that one model predicts context effects when the other does not, and vice versa. This was the strategy adopted in Experiment 3. In one distribution set, the ALs for the distributions being compared differed, but the proportion of comparisons strictly favoring stimuli from one distribution over the other was equated. If context effects are obtained, then they cannot be based on the distribution of responses (or rank information) but instead must be tied to other features of the stimulus distributions. In a second set of distributions, the ALs for the distributions were equated, but the RPs differed. If context effects are obtained, then they must be tied to the distribution of responses or rank infor-

mation. Evidence for this type of response-based effect was reported by Parducci and Haugen (1967), who used a similar manipulation in conjunction with the method of constant stimuli. They found that the PSE for lifted weights differed strongly for positively and negatively skewed distributions of comparison stimuli even when ALs for those distributions were roughly equated.

Although Experiment 3 was constructed as a critical test between the two models, it was possible that context effects would occur for both sets of distributions. Such a result would favor contributions to the contrast effect from both stimulus- and response-based processes. In such an event, the time course of these effects could be studied by examining the magnitudes of contextual effects as a function of response latencies. The expectation would be that stimulus-based effects would occur earlier than response-based effects.

Method

The method of Experiment 3 was nearly identical to that of Experiment 2 except for the stimulus sets that were paired together. The number of participants was also increased to allow a more careful examination of correlations between response latencies and the magnitude of the context effects.

Participants. Participants were 177 undergraduates from the University of South Carolina who received course credit for their participation. As in Experiment 2, a 2 × 2 × 36 factorial design was used. The between-subjects variables were distribution set (high AL [HA] vs. low AL [LA] or high-range positive skew [HP] vs. low-range negative skew [LN]) and distribution arrangement (HA or HP distributions on the left of the screen or on the right of the screen). Students were randomly assigned to one of the four between-subjects conditions.

Stimuli. As shown in Table 1, two square sizes (9 and 11 or 10 and 11) were common to each pair of distributions used in Experiment 3. The four additional stimuli making up the HA and LA distributions were chosen to create a difference in ALs but to equate the RPs. The four additional stimuli in the HP and LN distributions were chosen to equate the ALs but to created different RPs.

Results

Choice proportions. Choice proportions for Experiment 3 are shown in Table 6. The comparison of squares from LA and HA distributions provides a test of a stimulus basis for context effects. The AL model predicted context effects in these conditions, but the response-equalization model did not. The zero-step comparisons and the one-step comparison both indicated the presence of context effects. One-tailed z tests of one proportion indicated that the contrast effects on the 10-10 pair and the 11-11 pair were significant ($p < .05$). A one-tailed t test of the superadditivity of the one-step comparisons was also significant ($p < .05$). A 2 × 2 analysis of variance (ANOVA) on the zero-step comparisons revealed one significant result, a main effect of distribution arrangement, $F(1, 88) = 6.96, p < .05, MSE = 0.084$, reflecting greater effects of context when the HA distribution was presented on the left. Overall, the pattern of

Table 6
Proportions of Judgment Indicating the Target Square From the LA or HP Distribution Higher: Experiment 3

Stimulus pair[a]	Distribution pairing			Stimulus pair[b]	Distribution pairing		
	LA-HA	HA-LA	M		HP-LN	LN-HP	M
10-10	.593	.730	.662	9-9	.671	.761	.716
11-11	.626	.707	.667	11-11	.647	.746	.697
11-10	.981	.970	.976	11-9	.977	.977	.979
10-11	.144	.144	.144	9-11	.008	.008	.008

Note. LA = low adaptation level; HA = high adaptation level; HP = high-range positive skew; LN = low-range negative skew. [a] The first number is the square size from the LA distribution, and the second number is the square size from the HA distribution. [b] The first number is the square size from the HP distribution, and the second number is the square size from the LN distribution.

results supported the AL model predictions and tied contrast effects to properties of the stimulus distribution.

Comparison of squares from the HP and LN distributions provides a test of a response basis for context effects. The response-equalization model predicted context effects in these conditions, but the AL model did not. The zero-step comparisons indicated the presence of context effects. One-tailed z tests of one proportion revealed that the contrast effects on the 9-9 pair and the 11-11 pair were significant ($p < .05$). The two-step comparison did not exhibit context effects, which is consistent with other two-step comparisons presented in Experiment 1. A 2 × 2 ANOVA on the zero-step comparisons revealed one significant result, a main effect of distribution arrangement, $F(1, 85) = 5.55, p < .05$, $MSE = 0.071$, reflecting greater effects of context when the LN distribution was presented on the left. Overall, the pattern of results supported the response-equalization model predictions and tied contrast effects to properties of the distribution of responses.[5]

Relationship between context effects and response latency. The analyses of choice proportions provided support for the operation of both stimulus-based and response-based processes. Given that both processes affect paired comparisons, one may ask whether the time course of these processes is similar or different. Assuming that stimulus-based processes reflect a process of stimulus sampling across time, it is reasonable to assume that early samples may be most affected by distributional features extraneous to the stimuli being compared. As sampling increases, the information sampled will be tied directly to the stimuli being compared, and the effects of distribution should decrease. On the other hand, response-based context effects may be relatively unaffected by time or may actually increase with time if the response-based processes occur at a later stage of processing.

As a means of analyzing these possible effects, a multiple regression analysis was conducted in which the mean of the zero-order proportions was regressed onto the mean time across all trials and the distribution set. Including the interaction term in the equation significantly incremented the R^2 value, $F(1, 173) = 17.39, p < .001$, indicating that the

relationship between choice proportions and time was moderated by the distribution set variable. Separate regressions of choice proportions on response latencies were run for the two conditions; the results of these analyses are presented in Figure 1. For the HA-LA condition, the magnitude of the context effect decreased significantly with response latency ($r = -.37$), $F(1, 88) = 13.76, p < .001$. Including a quadratic term did not significantly increase the R^2 value ($p > .05$). For the HP-LN condition, the magnitude of the context effect increased significantly as a function of response latency ($r = .23$), $F(1, 85) = 4.75, p < .05$. Including a quadratic term did not significantly increase the R^2 value ($p > .05$). The differential time course of these effects provides further support that qualitatively different processes are responsible for contrast effects on paired comparisons.

Model analyses. The demonstration of context effects for manipulations of ALs when RPs were equated and for manipulations of RPs when ALs were equated implies that none of the three models described by Equations 6, 11, and 12 can account for the full pattern of data. However, if one assumes that the rank-dependent valuation effects examined in Experiment 2 are minimal, the data may be well represented by a mixed model in which both differences in ALs and differences in RPs contribute to contrast effects. Thus, Equations 11 and 12 can be combined into a single equation, as follows:

$$p(x_{i1} > x_{j2}) =$$

$$1/(1 + e^{-b[S_i - S_j + z_1(AL_2 - AL_1) + z_2(RP_{21} - RP_{12})]}), \quad (13)$$

where z_1 represents the weighting of the stimulus-based AL process and z_2 represents the weighting of the response-equalization process. When b is constant across conditions and each square is compared with itself, the logistic transformation of probability values can be described by the following linear equation:

$$\text{Logit}(p) = b + z_1(AL_2 - AL_1) + z_2(RP_{21} - RP_{12}), \quad (14)$$

where $\text{Logit}(p)$ is the natural logarithm of $p/(1 - p)$.

The assumption of a constant value of b seems plausible for the simultaneous presentation conditions. Thus, one convenient way to examine the predictive power of the models is to conduct a least squares linear regression of the logits for the zero-step proportions from Experiments 1–3 onto the parameters from the various models (e.g., as illustrated in Equation 14). The corresponding R^2 values were .806 for the AL model, .560 for the rank-dependent valuation model, and .783 for the response-equalization model. The combined model illustrated in Equation 14 yielded an R^2 value of .921, with coefficients for both the AL and the RP parameters significant ($p < .05$).

The fit of this model is illustrated in Figure 2. As can be seen, the most deviant point is for an extreme logit corresponding to a proportion of .988 from Experiment 1. Such

---

[5] Repeated measures MANOVAs based on the arcsine transformation yielded the same pattern of significance.
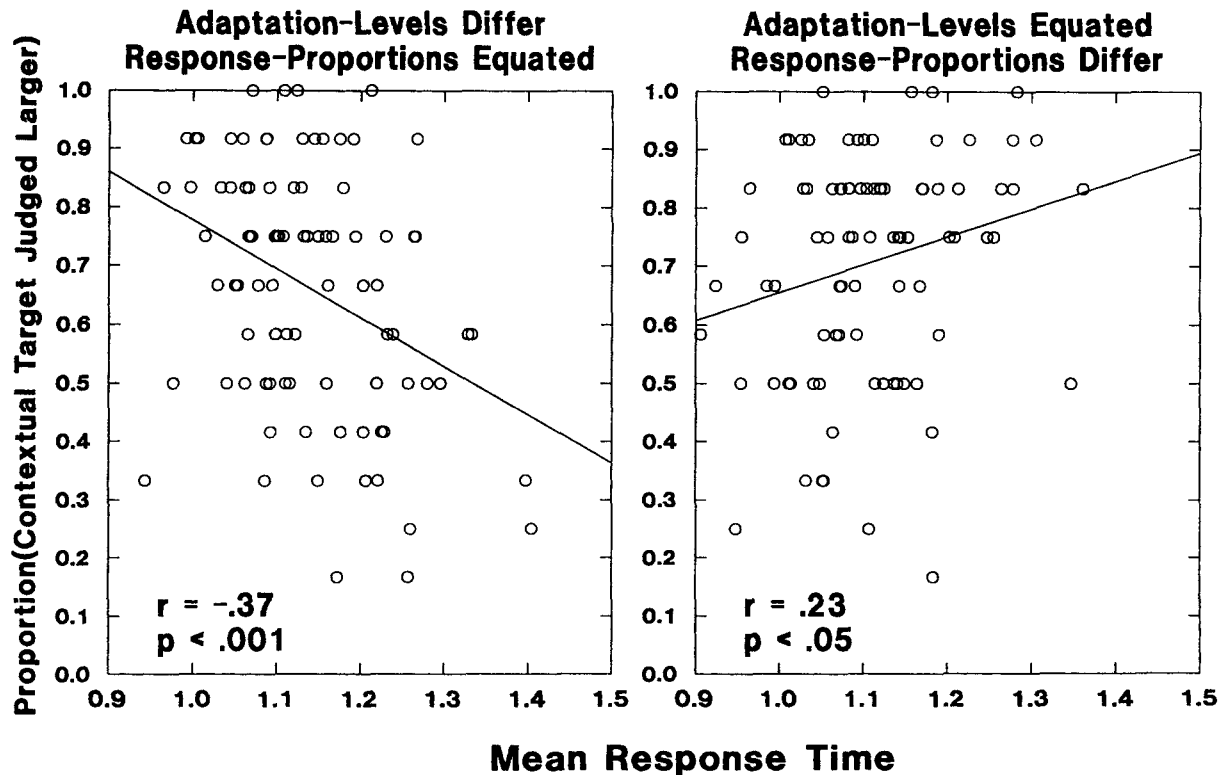
Figure 1.   Relationship between the magnitude of contextual effects and mean response time for the two sets of distributions used in Experiment 3. The opposite relationship under the two conditions provides support for an early occurring stimulus-based process and a later occurring response-based process.

extreme proportions require a very large number of observations to be reliably determined, and so this deviation is not particularly serious. The predicted data points for the conditions of Experiment 2 take on only two values, because both the response-equalization model and the AL model predict constant effects of context. Including parameters for the rank-dependent valuation model primarily affected the fit for these points, improving it only for the positive–negative skew conditions. Given the inconsistency of rank effects in Experiment 1 and the lack of a significant increment in the R^2 value associated with including this parameter, the data seem parsimoniously described by the two-factor model.

General Discussion

The combined results from the three experiments provided strong support for both a stimulus and a response basis of contrast effects in paired comparisons. Previous researchers have postulated both types of mechanisms, but they have not provided evidence to clearly distinguish between the two. The results from Experiment 3 provided the most clear-cut evidence by manipulating distribution means and paired RPs separately. The occurrence of significant context effects when the ALs differed but the RPs were

equated provided evidence for a stimulus-based process. The occurrence of significant context effects when ALs were equated but the RPs differed provided evidence for a response-based process.

It should be noted that the test of the two models presented in Experiment 3 depended critically on the independent manipulation of ALs and RPs. The calculation of AL was based on the assumption that scale values were linear to square widths, an assumption that has been experimentally verified for category rating tasks (Parducci & Wedell, 1986). However, scale values may be related to square width in a nonlinear fashion for other tasks, such as paired comparisons. One possibility is a logarithmic function, which would be consistent with an approximately equal discriminability scale inferred for square judgment (Parducci & Perrett, 1971). Another possibility is a power function. Wedell (in press) found that inferred scale values for squares in a pairwise dissimilarity judgment task were a negatively accelerated power function of square widths, with the power exponent varying from .25 to .60. What is common to these alternative functions is that they are all negatively accelerated on square width. As it turns out, a logarithmic function or any negatively accelerated power function, when applied to the critical conditions of Experiment 3, produces differences in ALs that predict effects in
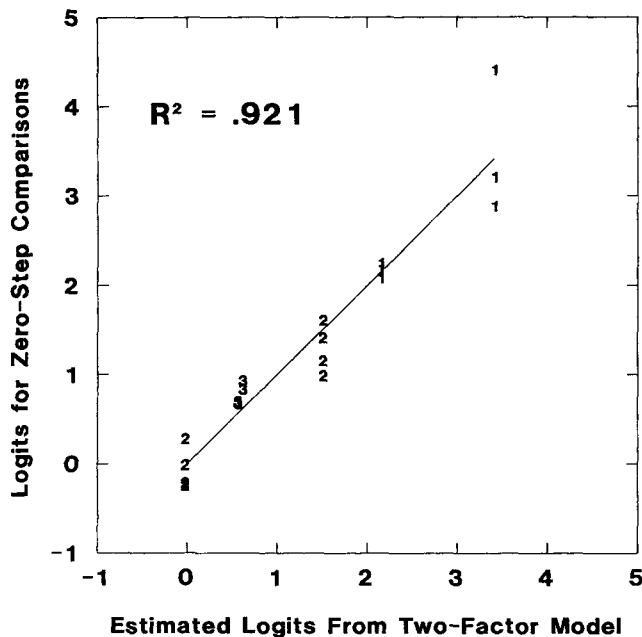
Figure 2.   Prediction of choice proportions for zero-step comparisons of a stimulus with itself from a two-factor model that weights both differences in adaptation levels and differences in response proportions. Scales are based on a logistic transformation of proportions.

the opposite direction than were obtained. Thus, even for these sets of plausible alternative psychophysical functions, the independent manipulation of RPs and ALs holds.

Time Course Dissociation of Contextual Processes

Perhaps the most convincing evidence for a two-process theory of context effects on paired comparisons is the dissociation between response latencies and context effects for conditions designed to isolate the different processes. When ALs were manipulated in Experiment 3, context effects were negatively associated with response latency. This relationship is consistent with a stimulus sampling conception of the choice process in which initial samples are strongly dependent on the stimulus distribution. However, as sampling continues across time, samples that are gathered are more directly linked to the stimuli being compared and thus become less context dependent.

When ALs were equated and RPs manipulated, context effects were positively associated with response latency. One possible explanation of this effect is that the response bias operates at a later stage in the judgment process, so that those who wait longer to respond will tend to exhibit a greater bias. A more parsimonious explanation is that the positive correlation was simply the result of the reduced effects of the stimulus-based process with time. Because ALs were equated in this condition, those responding early on the basis of differences in AL should show little effects of context. If differences in AL receive less weight as time

increases, then participants should show greater relative weighting of the response-equalization tendency.

Whatever the interpretation of the differential time course of the two processes, this dissociation provides one explanation of why differences in AL increased context effects in the simultaneous but not successive presentation conditions of Experiment 1. Introducing a 2-s delay between presentations of members of the stimulus pair added to the judgment time from the onset of the first square. With increased time, stimulus-based effects decreased, and so participants became insensitive to the manipulation of ALs. Because successive presentation is used for many paired-comparisons procedures, this explanation would suggest that effects of response equalization may be more pervasive than those of the stimulus-based AL process. Additional research is needed to determine the relative weighting of these processes for successive presentation conditions.

Rank-Dependent Processing

Experiment 2 was designed to test predictions that differentiate the rank-dependent valuation model from the other two models. The results from the bell–U comparisons provided no support for the model; the results from the positive–negative comparisons provided only weak support for the model. The lack of support for the rank-dependent valuation model is consistent with results from Mellers and Birnbaum (1982) in which graded difference judgments of lightness or darkness of dot patterns showed no rank dependency. Wedell (in press) also found a lack of rank dependency for dissimilarity judgments of dot patterns presented simultaneously in pairs. However, when either a delay between members of each pair was introduced or dot patterns were rated before dissimilarity judgments, rank dependency emerged. Wedell argued that rank information was likely to affect comparisons when either stimuli had to be held in memory or rank information was encoded in prior categorical associates.

Wedell (in press) found a similar pattern of results for squares, except that there were small but significant rank-dependent effects even when the pair of squares was presented simultaneously with no prior category ratings. These effects were much larger when a delay was introduced between presentation of members of the pair. One avenue for future research is to reconcile the lack of rank-dependent effects in paired comparisons of simultaneously presented squares found in Experiment 2 with the significant rank-dependent effects reported for simultaneously present squares by Wedell (in press). One possible explanation is that the dissimilarity rating task simply provided a more powerful test of the effects of manipulating rank. Another possibility is that the process of making a graded judgment is more likely to involve rank-dependent processes than the process of making a simple dichotomous judgment. Finally, future tests of the rank-dependent valuation model in paired comparisons should examine successive presentation conditions. Wedell (in press) found that these delay conditions produced the largest effects of rank manipulations on dissimilarity ratings.

Loci of Context Effects

The loci of context effects have long represented a subject of controversy in the psychological literature. There now seems to be substantial evidence that the processes producing context effects can be effortful or automatic, can be stimulus based, and can occur early or late in the sequence of processing (Wedell, 1990, 1994). The present set of experiments argues for an early stimulus-based process and a later response-based process operating on paired comparisons. The issue of automatic versus controlled processing was not addressed in this work, but the generally fast response latencies suggest that the process may well be automated.

The use of two or more contextual sets within the same judgment task provides a way to directly assess disordinal effects of context on judgment. In addition to paired comparisons, Marks and his colleagues (Marks, 1991, 1992; Marks & Warner, 1991; Rankin & Marks, 1991, 1992) have found similar effects using the dual distribution paradigm in magnitude estimation tasks. Their experiments have demonstrated that this type of contrast effect depends on the qualitative similarity of the stimuli making up the distribution sets as well as a number of other features of the stimuli. For example, Marks (1991) reported context effects on judgments of loudness for sets that differed in pitch but no context effects on judgments of pitch for sets that differed in loudness. A fuller theory of context effects must account for the dependence of these effects on properties of the stimulus dimension being judged. Such a theory will probably include multiple contextual processes that act at different levels and on different structures within the cognitive processing system.

References

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

Davidson, D., Suppes, P., & Seigel, S. (1957). *Decision-making: An experimental approach.* Stanford, CA: Stanford University Press.

Erlebacher, A., & Sekuler, R. (1971). Response equalization: A bias model for psychophysics. *Perception & Psychophysics, 9*, 315–320.

Fechner, G. T. (1966). *Elements of psychophysics: Vol. 1* (E. G. Boring & D. H. Howes, Eds.; H. E. Adler, Trans.). New York: Holt, Rinehart & Winston. (Original work published 1860)

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

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

Helson, H., Michels, W. C., & Sturgeon, A. (1954). The use of comparative rating scales for the evaluation of psychophysical data. *American Journal of Psychology, 67*, 321–326.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–293.

Link, S. W. (1992). *The wave theory of difference and similarity.* Hillsdale, NJ: Erlbaum.

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

Manis, M., Nelson, T. E., & Shedler, J. (1988). Stereotypes and social judgment: Extremity, assimilation, and contrast. *Journal of Personality and Social Psychology, 55*, 28–36.

Marks, L. E. (1991). The slippery context effect in psychophysics: Intensive, extensive, and qualitative continua. *Perception & Psychophysics, 51*, 187–198.

Marks, L. E. (1992). The contingency of perceptual processing: Context modifies equal-loudness relations. *Psychological Science, 3*, 285–291.

Marks, L. E., & Warner, E. (1991). Slippery context effect and critical bands. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 986–996.

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

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

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

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). Berlin: VEB Deutsher Verlag der Wissenschaften.

Parducci, A., Calfee, R. C., Marshall, L. M., & Davidson, L. P. (1960). Context effects in judgment: Adaptation level as a function of the mean, midpoint, and median of the stimuli. *Journal of Experimental Psychology, 60*, 65–77.

Parducci, A., & Haugen, R. (1967). The frequency principle for comparative judgments. *Perception & Psychophysics, 2*, 81–82.

Parducci, A., Marshall, L. M., & Degner, M. (1966). Interference with memory for lifted weight. *Perception & Psychophysics, 1*, 83–86.

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

Parducci, A., & Sandusky, A. J. (1970). Limits on the applicability of signal detection theories. *Perception & Psychophysics, 7*, 63–64.

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.

Rankin, K. M., & Marks, L. E. (1991). Differential context effects in taste perception. *Chemical Senses, 16*, 617–629.

Rankin, K. M., & Marks, L. E. (1992). Effects of context on sweet and bitter tastes: Unrelated to sensitivity to PROP (6-n-propylthiouracil). *Perception & Psychophysics, 52*, 479–486.

Restle, F., & Levison, M. (1971). Method of constant stimuli: Invalidity to the third power. *Perception & Psychophysics, 9*, 312–314.

Schneider, B., & Parker, S. (1990). Does stimulus context affect loudness or only loudness judgment? *Perception & Psychophysics, 48*, 409–418.

Seigel, S. (1957). Level of aspiration and decision making. *Psychological Review, 64*, 253–262.

Thurstone, L. L. (1927a). A law of comparative judgment. *Psychological Review, 34*, 273–286.

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

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review, 76*, 31–48.

Wedell, D. H. (1990). Methods for determining the locus of context effects in judgment. In J. P. Caverni, J. M. Fabre, & M. Gonzalez (Eds.), *Cognitive biases* (pp. 285–302). New York: Elsevier Science.

Wedell, D. H. (1991). Distinguishing among models of contextually induced preference reversals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 767–778.

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. (in press). A constructive-associative model of the contextual dependence of unidimensional similarity. *Journal of Experimental Psychology: Human Perception and Performance.*

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.

1995 Research Awards in Experimental Psychology

The American Psychological Association's Division of Experimental Psychology (Division 3) announces the following awards, presented at the Annual Convention in New York in August 1995. The awardees were determined by review of the research submitted to or published in the APA's Journals of Experimental Psychology in the past year by relatively new investigators. The intention is to provide early recognition for new scholars whose research contributions are especially promising.

Hilary A. Broadbent
Young Investigator Award
Animal Behavior Processes

Daniel Read
Young Investigator Award
Applied Psychology

Robert Goldstone
Young Investigator Award
General Psychology

Steven Lindsay
Young Investigator Award
Human Perception and Performance

Robert Goldstone
Young Investigator Award
Learning, Memory, and Cognition