

Range effects of an irrelevant dimension on classification

SCOTT A. HUETTEL and GREGORY R. LOCKHEAD
Duke University, Durham, North Carolina

In univariate classification tasks, subjects sort stimuli on the basis of the only attribute that varies. In orthogonal classification tasks, often called *filtering* tasks, there additionally are trial-to-trial variations in irrelevant attributes that the subjects are instructed to ignore. Performance is generally slower in filtering tasks than in univariate control tasks. We investigated this slowing in experiments of how the range of irrelevant trial-to-trial variation affects responses in pitch/loudness classification tasks. Using two levels of pitch and of loudness as stimuli, Experiment 1 replicated prior work showing that responses are slowed more when the range of the irrelevant dimension is made larger. Also in Experiment 1, sequential analyses showed that response time depends both on sequence and on the stimulus set independent of sequence. Experiments 2 and 3 used several levels on the irrelevant dimension and showed that responses to categorize loudness are slowed more by larger trial-to-trial pitch differences, but only on trials when the response repeats. When the response changes, performance is essentially unaffected by trial-to-trial irrelevant variation. This interaction supports the conclusion that slowed average performance in orthogonal classification tasks, which is known as *Garner interference*, is not due to difficulties that subjects have in filtering stimulus attributes. It is due to how subjects process successive stimulus differences. We call for more frequent reports of sequential analyses, because these can reveal information that is not available from data averages.

Often, people cannot ignore irrelevant information when they classify a stimulus attribute. For example, when people are asked to classify auditory tones according to pitch, their performance is worse (slower and with more errors) when the timbre of the tone varies randomly between trials than if the timbre remains fixed between trials (Crowder, 1989). Irrelevant variation in loudness similarly interferes with performance when tones are classified according to pitch (Lockhead, 1992b; Melara & Mounts, 1994).

This result—the fact that irrelevant variation on one dimension interferes with classification on another dimension—is called *Garner interference* (Garner, 1974). Whenever there is such interference, the stimulus dimensions interact, which is contrary to what should occur if people could abstract a feature from a stimulus and process it independently of other features, as Stevens (1975) proposed. There have been many attempts to understand this apparent interference (Garner, 1970; Lockhead, 1966, 1992b; Melara & Marks, 1990; Melara & O'Brien, 1987; Pomerantz & Garner, 1973; Shepard, 1964), none of which was fully successful. Each of these attempts addressed characteristics of individual stimuli, and all of these experimental approaches compared average perfor-

mance in one condition to average performance in another condition.

A factor that cannot be addressed by such studies of averaged data is any effect of the sequential structure of the stimuli, because sequential information is lost when data are averaged across trials. Sequence should be considered for at least two reasons: stimulus sequences are different in tasks that have irrelevant stimulus variation than in tasks where there is no such variation, and, as Felfoldy (1974) showed, performance depends on sequence. Even so, Garner interference essentially has not been examined for sequence effects beyond Felfoldy's study. Some such examinations are reported here.

Another factor that is not examined in Garner tasks but should be studied is how the range of variation on the irrelevant dimension affects performance. Range affects performance in absolute judgment tasks when stimuli differ on only one dimension (Durlach & Braida, 1969; Gravetter & Lockhead, 1973; Parducci, 1965; Pollack, 1953) and in a bivariate Garner task when the range of the irrelevant dimension is increased (Lockhead, 1992b; Melara & Mounts, 1994). In each case, performance is poorer when range is larger. One might expect sequence effects and the range of the irrelevant dimension to be related. This is because performance is poor when the irrelevant range is large and because successive differences between stimuli within a task are large, on average, when the range is large. This fact allows the suggestion that both range effects and Garner interference are associated with trial-to-trial sequence effects. To consider this, effects of both stimulus range and stimulus sequence are examined.

This research was partially supported by a National Science Foundation fellowship to S.A.H. Correspondence should be addressed to S. A. Huettel, Department of Psychology, Duke University, Durham, NC 27708 (e-mail: huettel@psych.duke.edu).

—Accepted by previous editor, Myron L. Braunstein

Theoretically, one reason for examining sequence and range effects in Garner tasks is to pursue a proposal by Lockhead and King (1983) that describes processes that seem to be involved when people classify stimuli in magnitude estimation and absolute judgment tasks. The proposal for these univariate tasks is that people judge a stimulus by comparing it with their memory of the prior stimulus, and this comparison both requires less time and is more precise when successive stimuli are more similar; the comparison task is then easier. This thesis describes much of the univariate judgment data. The experiments reported in this paper examined whether the thesis also describes performance in Garner tasks.

Range and sequence effects are appropriate empirical measures to use to examine this trial-to-trial comparative judgment thesis. This is because it states that performance is less precise when successive stimuli are more different. Successive stimuli are more different, on average, when the level of the irrelevant dimension changes between trials than when it remains fixed, and the magnitude of this change will be greater when the range of irrelevant variation is greater. Hence, the predictions are that performance between tasks is poorer in conditions with a large stimulus range than in conditions with a small stimulus range and that performance within tasks is poorer on trials when successive stimuli are physically more different.

However, Lockhead and King (1983) cannot predict, without additional assumptions, how effects of range and sequence interact across relevant and irrelevant dimensions. Are range and sequence effects due to the entire stimulus, due to only the attended dimension, which would mean the irrelevant dimension is not involved, due to only the irrelevant dimension, which would implicate some attention process, or due to some combination of these? The experiments reported here addressed these questions, as well as questions about effects of range, both within and between conditions.

To anticipate the work detailed ahead, we found that increased range of the irrelevant dimension results in decreased quality of performance on the relevant dimension. We also found that range interacts with the subject's processing of the stimulus dimensions. It appears that subjects compare successive stimuli holistically. This allows a fast response when the stimulus repeats (repeat the last response). This comparison also allows a fast response when the two stimuli are very different (press the other key from the prior trial). However, such a comparison results in slowed performance when successive stimuli are very different, because only the irrelevant dimension value changes between trials, whereas the relevant dimension value repeats. Here, the stimulus change signals a response change, but this response is wrong and must be halted.

We shall conclude that subjects essentially perform a *same-different* task between the current and previous stimuli in these classification tasks, even though the subjects' instruction is to identify the level of the relevant dimension for each stimulus. We suggest that conflict

from these two response tendencies results when only the irrelevant value changes between trials, and this is the primary source of Garner interference.

Classification Tasks

In classification tasks, subjects sort stimuli into different categories (E. E. Smith & Medin, 1981). Some tasks, such as absolute identification, contain stimuli that vary only on the dimension that is judged, and subjects classify each stimulus according to its value on that dimension (e.g., according to loudness). Commonly, there are 10 category levels on the judged dimension, but any number of levels are possible (cf. Miller, 1956). Tasks that involve variation on only one stimulus dimension are called *univariate tasks*.

Other tasks, as focused on here, designate one stimulus dimension as relevant, with subjects discriminating two values on that dimension, whereas stimulus values on an irrelevant dimension vary randomly from trial to trial. One set of such tasks, known as *Garner tasks*, examines effects on performance of various relations between relevant and irrelevant stimulus dimensions (Garner, 1974; Pomerantz, 1991).

The Garner stimulus set uses two values on each of two dimensions, and there are three basic tasks. In baseline tasks, which use only two stimuli, the subject discriminates between the two objects that differ in only a single dimension. Stimuli are presented successively, one at a time and in random order, for many trials. These tasks provide a performance measure for judgments when there is no variation from trial to trial in irrelevant information. In correlated tasks, the two levels of the two dimensions covary so that the subject might use either dimension to guide responses. Again, there are only two stimuli in a condition, presented one at a time and randomly on successive trials. In filtering (or orthogonal) tasks, any of the four stimuli in the set may occur randomly from trial to trial. These tasks again require the subject to classify objects on a relevant dimension, but now there is random trial-to-trial variation between two levels of an irrelevant dimension. We employed only baseline and filtering tasks in the present experiments.

One possible outcome from the filtering task is that subjects can ignore the characteristics of the irrelevant dimension, such that judgment accuracy and response time (RT) are independent of its nature. When this occurs, as it often does, the processing of one component is considered to be separate from the processing of the other, and the dimensional pair is called *separable* or *analyzable*. Another outcome that also occurs often is that random variation on the irrelevant dimension affects performance. Generally, performance is then best (fastest and with fewest errors) in the correlated tasks, poorer in the baseline tasks, and poorest in the filtering tasks. Dimensional pairs with this relation are considered to be interacting (Melara & Marks, 1990). Dimensions that interact due apparently to initial holistic processing, such as the color dimensions of hue, saturation, and lightness, are called

integral (Garner, 1974; Lockhead, 1966, 1972). Some interacting dimensions are nonintegral and possess some cross-modal equivalencies between their components; an example is auditory pitch combined with vertical position (Marks, 1987; Melara & O'Brien, 1987). Finally, some configural pairs interact due to the geometric structure of the components (Pomerantz, 1991). No independent basis exists for determining if and how dimensions interact, and so empirical studies must be conducted in each case. The dimensions used in the present experiments, auditory pitch and loudness, are known to be integral (Grau & Kemler-Nelson, 1988; Melara & Marks, 1990), and so interacting effects of irrelevant dimension variation were expected.

A robust indicator of an interacting relation between dimensions is the presence of Garner interference. This is defined by a positive difference between the mean RT in filtering tasks, where an irrelevant dimension varies in value from trial to trial, and the mean RT in baseline tasks, where the value of the irrelevant dimension remains fixed between trials. When there is Garner interference, the filtering task requires more time (Pomerantz, 1991; but see van Leeuwen & Bakker, 1995), and the source of such interference has been described as a failure of selective attention to the relevant dimension (e.g., Melara & Mounds, 1994; Pomerantz, Pristach, & Carson, 1989). Such failure might occur because of processing of the stimulus itself (Broadbent, 1971), because of overall effects of the total stimulus set on processing (Garner, 1962), because of memories of preceding individual stimuli (Lockhead, 1992b), or because of some combination of these.

Range Effects

Univariate tasks. Stimulus range affects performance in all classification tasks where range has been examined. Responses are more variable and require more time in tasks with larger total stimulus ranges, even when the number of stimuli and the discriminability between adjacent stimuli in the sets are similar (Durlach & Braida, 1969; Parducci, 1965; Pollack, 1953). For example, consider when the task is to absolutely identify three 1000-Hz tones that differ only in loudness. The tones are played, one at a time in random order for many trials, and the subject is instructed to classify each tone by pressing the appropriate one of three keys. When the tone intensities are 70, 72, and 74 dB, the quieter two members (70 and 72 dB) are identified more accurately than when the intensities are 70, 72, and 80 dB, even though the same values are judged in both instances (Gravetter & Lockhead, 1973). That is, accuracy decreases with increasing stimulus range. Accordingly, these 70- and 72-dB tones are even more difficult to identify when the third tone is 85 or 60 dB. This range effect is not restricted to people as subjects; the seemingly identical result is found when pigeons classify lights according to flicker rate (Hinson & Lockhead, 1986). It seems generally to be the case that the range of the stimulus set affects performance in iden-

tifying individual members of the set, with better performance (greater accuracy and shorter RTs) for smaller ranges.

Similar results are found in magnitude estimation studies in which subjects assign numbers to stimuli such that the ratio of successive response numbers is equivalent to the perceived ratio of successive stimulus intensities (Stevens, 1975). For example, if the subject's response to the loudness of a tone is 50, and if the next tone sounds twice as loud, then it should be called 100. If subjects actually do absolutely identify the loudness of each tone independent of context, and if they follow instructions, then the same two tones played in succession should have the same apparent relation, independent of the range of other tones in the set. But they do not. The reported ratio between the tones decreases when the range of tones in the experiment is made larger (Teghtsoonian, 1973). It is generally the case, for sets of univariate stimuli, that increased stimulus range results in poorer discrimination between fixed members of the sets.

Bivariate tasks. While range is well known to affect performance in univariate tasks, range effects have not been as well studied in classification tasks in which there is variation on two or more dimensions. What has been well studied is Garner interference (the effect of the presence or absence of any variation on a second dimension). Some data indicate that the magnitude of this interference increases with increased range of variation in the irrelevant dimension across tasks. For pitch and loudness, overall RT is slower with larger ranges of irrelevant variation (Lockhead, 1992b; Melara & Mounds, 1994). Particularly, overall RT is fastest in baseline tasks, where there is no irrelevant variation, slower in filtering tasks with a narrow range of irrelevant variation, and slowest in filtering tasks with a wide range of irrelevant variation. These results extend to judgments of auditory pitch in the presence of irrelevant variation in the vertical position of a line (i.e., corresponding stimulus dimensions; Ben-Artzi & Marks, 1995). It may be generally the case that a larger irrelevant range has a greater effect on performance when people judge values of stimulus attributes.

Sequence Effects

Univariate tasks. Judgments in univariate tasks, both absolute identification and magnitude estimation, depend on stimulus sequence and on range. Independent of the number of stimuli to be classified, responses are not only more variable in conditions where successive stimuli are more different on average (Baird, Berglund, Berglund, & Lindberg, 1991; Gravetter & Lockhead, 1973; Luce & Green, 1978) but they are also more variable on trials when the prior stimulus was more different (Holland & Lockhead, 1968). Responses to each stimulus tend toward (i.e., assimilate toward) the value of the stimulus or the response on the prior trial. Furthermore, the amount of this shift in the response is greater when the difference between successive stimuli is greater (Lockhead & King, 1983). These sequence effects, measured within condi-

tions, must be correlated with across-conditions range effects because, on average, when the overall stimulus range is large, then the average difference between successive stimuli is also large.

Bivariate tasks. There are other sequence effects, in addition to the increased response variability and response assimilation just referenced, that are affected by repetitions of the stimulus and of the response. Performance is often faster and more accurate on trials when the stimulus repeats from the preceding trial than on nonrepetition trials, for both univariate and bivariate tasks (Bertelson, 1961, 1963; Hyman, 1953). Also, trials on which the response repeats are often faster than trials on which the response changes in bivariate tasks (Bertelson, 1965; Eichelman, 1970; Felfoldy, 1974; Kornblum, 1973; E. E. Smith, Chase, & P. G. Smith, 1973). However, other work failed to identify systematic facilitation associated with response repetition when the stimulus does not repeat (Ells & Gotts, 1977; Hinrichs & Krainz, 1970; Li & A. F. Smith, 1992; M. C. Smith, 1968; Williams, 1966). Fletcher (1981) concluded that the presence or absence of repetition effects depends on practice: Early in practice, subjects display a strong stimulus repetition effect, which progresses into a response repetition effect later in practice (see also Bertelson, 1965; Rabbitt, 1968). Pashler and Baylis (1991) report that response-repetition effects, as separate from stimulus-repetition effects, occur only when the category members differ on some superficial attribute (e.g., color, in a letter classification task; see also Lockhead, Gruenewald, & King, 1978). Several things need to be untangled for us to understand what occurs when people judge stimulus attributes; some of them are examined here.

Felfoldy (1974) reported sequence effects in baseline and filtering Garner tasks, using rectangles with two values each of height and width as stimuli. With height (or width) as the relevant dimension and width (or height) varying irrelevantly, he found both stimulus- and response-repetition effects. For both baseline and filtering tasks, responses were faster when the stimulus was preceded by itself than when it was preceded by any other stimulus, and responses were faster when the response repeated from the previous trial than when the response changed between trials. Felfoldy suggested that the noted overall RT difference between control and filtering tasks (Garner interference) may result from differences in sequential structure between the tasks. The explanation is this: The control task is composed of equal numbers of two types of trials, those where the stimulus and response both repeat (with faster RTs) and those where the stimulus and response both change (with slower RTs). For the filtering task, only one fourth of the trials are the fast stimulus and response repetitions, another fourth are the slower stimulus-change/response-repeat trials, and half are the even slower stimulus-change/response-change trials. Therefore, if RT is similar for a particular trial type across tasks, the filtering task has a higher proportion of the slower trial types, and this could account for the slower performance overall.

Fletcher (1981) conducted a similar two-dimension classification task, but with letter identity as the relevant dimension and with type of mask (two levels of amount of occluding dots) as the irrelevant dimension. This relevant dimension, alphabetic letter, is not dimensional in a continuous sense because the different possibilities for each attribute are composed of different features, rather than of different values on the same dimension (Garner, 1978). Furthermore, the irrelevant dots partially masked the relevant letters, providing what Garner (1970) calls a *state limitation* (the critical information does not always enter the system), and this, rather than a process limitation, may account for why performance was interfered with by the irrelevant dots. Yet, the stimulus features appear to interact, in that subjects were unable to completely filter out the irrelevant information. Early in practice in Fletcher's study, responses were fastest to stimulus repetitions, but, late in practice, the irrelevant dimension did not affect processing. RTs for all response repetitions (regardless of irrelevant-dimension level) were faster than for any other sequence. Fletcher varied the degree of interference from the dot mask (by including both 5-dot and 10-dot conditions) but did not observe any systematic increase in the strength of these effects.

Crowder (1989) investigated sequence effects in a *same-different* task with variation on two stimulus dimensions. The stimulus set was composed of nine sounds, created by the factorial combination of three pitches, each separated by one musical note, played with three possible timbres (three different musical instruments). Subjects judged whether two successively played tones had the same or different pitches. For "same" judgments, when the relevant dimension level of pitch repeated, RTs were faster when the irrelevant dimension of timbre also repeated (737 msec) than when the irrelevant dimension changed (941 msec). But, for "different" judgments, RTs were similar (about 780 msec) regardless of whether the irrelevant dimension also changed. Crowder does not report any range effects, although his experimental design contains three levels of pitch. Dixon and Just (1978) do report range effects in *same-different* tasks. Subjects judged if ellipses that varied in width had the same height, and if colors that varied in tint had the same hue. "Same" responses, to same height and same hue pairs, were slowed more by larger irrelevant differences between the two stimulus members, but "different" responses did not depend on this difference.

Lockhead (1992b) conducted classification tasks similar to Crowder (1989) and Dixon and Just (1978), using pitch and loudness as the stimulus dimensions. His experiments used feature classification where stimuli are presented successively one at a time, rather than *same-different* judgments where two stimuli are directly compared. He found both range and sequence effects. Responses were faster when the range of the irrelevant dimension was smaller, and responses also faster when the irrelevant attribute repeated from the previous trial than when it changed from the previous trial. That study

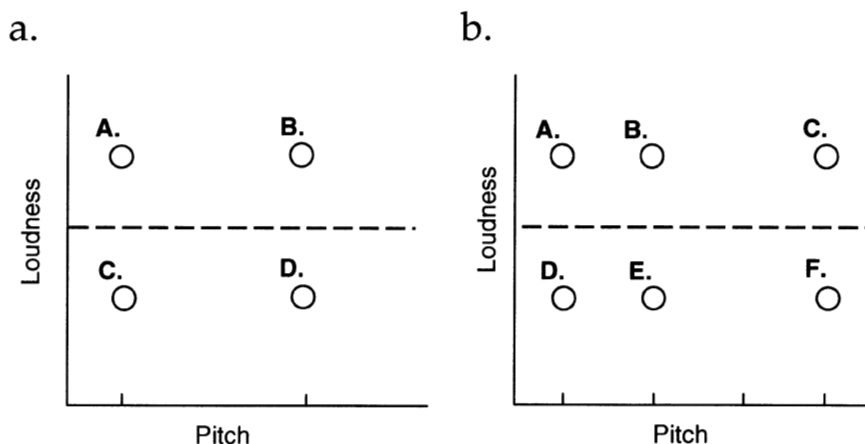


Figure 1. Sample stimulus sets of the form used in (a) Experiment 1A and (b) Experiment 2B. These are intended to complement the discussion of sequential analyses and terminology used in the current study (see text).

provided only a partial analysis of sequence effects in that it did not address whether irrelevant change has an additive effect to relevant change (cf. Felfoldy, 1974) or an interactive effect as a function of trial-to-trial relevant-dimension change (cf. Crowder, 1989) or whether responses interact with response type and irrelevant change, or some other effect.

Together, the above studies provide conflicting predictions for the pattern of sequence effects to be observed in the present experiments. Felfoldy (1974) and Fletcher (1981) predict that, for filtering tasks, RTs should be fastest to stimulus repetition trials, almost as fast or equally fast to response repetition trials, and slowest to response change trials. This pattern would be similar to that in univariate tasks, consistent with the suggestion of Lockhead and King (1983), where RTs are slower for larger trial-to-trial change between stimuli. In contrast, the results of Crowder (1989) suggest that changes in an irrelevant stimulus dimension slow response repetitions but do not affect response changes. Dixon and Just (1978) further suggest that this dependence on the irrelevant dimension is greater for larger irrelevant differences. In the present experiments, we analyzed RTs as a function of stimulus sequence, hoping to disambiguate these possible patterns for the dimensions of pitch and loudness. We investigated sequence effects both within a single range condition and across range conditions.

The Present Experiments

In Experiment 1, we addressed across-task range effects in orthogonal classification. The subjects judged a relevant dimension while an irrelevant stimulus dimension varied randomly from trial to trial. Keeping the two values of the relevant dimension constant across conditions, we manipulated the range of values on the irrelevant dimension. We examined effects of this range manipula-

tion on overall RT and on RT as a function of stimulus sequence.

In Experiments 2 and 3, we investigated range effects within a single orthogonal classification task. For this, we added values on the irrelevant dimension to the values used in Experiment 1. This changed the sequential structure within each condition by creating multiple magnitudes of irrelevant change between trials within each task, rather than having only one level of irrelevant change in the task. In Experiment 2, we added one irrelevant value; in Experiment 3, we added many values. Using the sequential effects obtained in Experiment 1 as a guide, we investigated whether RT is dependent on the magnitude of trial-to-trial irrelevant variation (i.e., a within-conditions range effect).

We measured response accuracy and RT as a function of stimulus sequence (one trial back only) when subjects classified tones that varied between trials in auditory pitch and loudness. Within each task, the stimulus set was composed of stimuli with different values of loudness and/or pitch. The subjects classified the stimuli according to the values, low or high, on a designated relevant dimension. The other dimension was irrelevant to the task. In every case, there were two possible stimulus values on the relevant dimension, one for each response category, but there were 1, 2, 3, 8, or 12 values on the irrelevant dimension, depending on the experimental condition.

Figure 1a shows a sample form of the stimulus set in Experiment 1. The axes represent the dimensions of pitch and loudness (acoustic frequency and amplitude), with the dashed line separating the stimuli along a relevant dimension of loudness as an example task. Each circle indicates a single stimulus (one pitch–loudness combination). Consider when Stimulus A was presented on Trial 1. Any of the four stimuli could have been presented on the next trial, so there were four possible stimulus sequences

for Trial 2. If Stimulus A was presented again, then the values repeated on both the relevant dimension and the irrelevant dimension. We label this sequence *R-R*; the first letter indicates the relevant dimension, which repeated (R) from the previous trial, and the second letter indicates the irrelevant dimension, which repeated (R) as well. If, instead, Stimulus B was presented on Trial 2, this was coded as the *R-C* category; the relevant-dimension value repeated (R), and the irrelevant-dimension value changed (C). Similarly, if Stimulus C was presented on Trial 2, it would be the *C-R* category; if Stimulus D was presented on Trial 2, it would be the *C-C* category. Any stimulus could fall into any category, depending on what preceded it.

Figure 1b provides an example of the type of stimulus set used in the later experiments. In Experiments 2 and 3, we adopted stimulus sets with more than two values on the irrelevant dimension, so that we could examine range effects within a single task. The specific sequence effects investigated are discussed at the beginning of those experiments.

EXPERIMENT 1 Irrelevant-Dimension Range Effects Across Tasks

Experiment 1 had two related purposes. First, we investigated across-conditions range effects on RT. If a range effect was present in the overall RT data (e.g., if wide range conditions were significantly slower than narrow range conditions), then we would analyze whether changes in the sequential structure across ranges accounted for this overall effect. Second, by recording RTs as a function of stimulus sequence, we determined a pattern of sequence effects for comparison with the within-conditions range effects (if any) of Experiments 2 and 3.

Increases in overall stimulus range slow RT in classification tasks for interacting dimensions (Ben-Artzi & Marks, 1995; Lockhead, 1992b; Melara & Mounts, 1994). To test for both set (Garner, 1962) and sequence effects of irrelevant-dimension range across tasks, we adopted the procedure of Lockhead (1992b). Stimuli were sine-wave tones varying in intensity and frequency. In each experimental condition, there were two values on a designated relevant dimension, so that the discrimination relevant to the task was the same throughout the experiment. In control conditions, there was no variation on the irrelevant dimension; these were the baseline tasks. In filtering conditions, the range between values on the irrelevant dimension was manipulated across conditions, from small but discriminable in the narrow conditions to large in the wide conditions.

To replicate the prior findings that range and trial-to-trial change in the irrelevant dimension affect performance (cf. Lockhead, 1992b), RTs should increase monotonically with increases in the range between values

on the irrelevant dimension, and responses should be faster when the irrelevant level repeats than when it changes. If this effect replicated successfully (and it did), we then planned to analyze RT as a function of both stimulus range and sequential structure. The purpose was to learn whether the overall range effect is completely due to sequential changes (i.e., Can the suggestion of Felfoldy, 1974, be extended to across-conditions range effects?).

Method

Subjects

Eight Duke University students participated in Experiment 1 (4 in Experiment 1A, and 4 in Experiment 1B). All subjects were paid \$8.

Stimuli

Stimuli were generated by a SoundBlaster-compatible soundcard on an IBM-compatible computer. The subjects heard the tones through Realistic Nova-43 headphones. All tones were 250 msec in duration with ramped onset and offset. Intensity calibrations were made using a Realistic sound-level meter (33-2055).

Tasks

In univariate tasks, the subjects discriminated between two stimuli differing only along the dimension of classification. For Experiment 1A (loudness judged), two univariate tasks measured whether these judgments depend on the particular frequencies used. The subjects judged the loudness of 1000-Hz tones played randomly across trials at 70 dB or 72 dB; in another condition, they judged 1500-Hz tones at 70 dB or 72 dB. For Experiment 1B (pitch judged), there was one univariate task, 70-dB tones played at 1000 Hz and 1015 Hz.

In filtering tasks, the subjects again classified stimuli according to their values on a single (relevant) dimension and were instructed to ignore variation on the other (irrelevant) dimension. For Experiment 1A, the intensities judged were 70 and 72 dB in three conditions of frequency range: *narrow* (15 Hz; 1000- and 1015-Hz stimuli), *medium* (45 Hz; 1000- and 1045-Hz tones), and *wide* (500 Hz; 1000- and 1500-Hz tones). For Experiment 1B, the frequencies were played at 1000 and 1015 Hz, in four conditions of loudness range: *narrow* (4 dB; 70- and 74-dB tones), *medium-narrow* (8 dB; 70- and 78-dB tones), *medium-wide* (12 dB; 70- and 82-dB tones), and *wide* (16 dB; 70- and 86-dB tones).

Procedure

All subjects were tested in an isolated, quiet room. Initial instructions advised the subjects of the classification dimension (pitch or loudness) and encouraged them to maximize the speed of their responses while maintaining a low error rate. Tasks and conditions were presented in separate blocks of trials, the order of which was counterbalanced across subjects.

At the beginning of each condition, the subjects were instructed to listen to the test tones as many times as desired by pressing an appropriate button. There was then a series of practice trials of randomly selected stimuli (250 msec in duration), with feedback, which ended when the subject was correct on a total of 25 trials. RT was measured from the stimulus onset to when the subject pressed one of the two designated keys on the keyboard. The interstimulus interval was 2 sec. Then, following a second display of the instructions, 100 test trials without feedback were presented. Each subject participated in 10 blocks (2 of each baseline and filtering condition) of 25+ practice and 100 test trials, with a 5-min break between the 5th and 6th blocks. Each session lasted approximately 1 h.

Table 1
Response Times (RTs; in Milliseconds) and Error Rates (% Errors) for Range Conditions in Experiments 1A (Loudness) and 1B (Pitch)

Range	RT	% Error
Loudness (Hz)		
1000 (Control)	558	5.5
1500 (Control)	542	4.0
15	582	5.0
45	596	6.0
500	621	10.0
Pitch (dB)		
Control	616	3.6
2	634	5.6
6	629	4.3
10	651	5.3
14	673	5.0

Results

Experiment 1A: Loudness Judgments

Errors. Error rates and RTs for the experimental conditions of Experiment 1A are presented in Table 1. The overall error rate was 6%. There was a positive correlation between RT and error rate ($r = .87, p < .05$). We did not analyze the errors as a function of stimulus sequence because of the small number of errors per sequence condition (generally 2–4 errors per condition per subject).

Response times. Only correct trials are included in the reported RT analyses. According to planned *t* tests, responses in each control condition were faster than in

any orthogonal condition, and not significantly different from each other (for every test, $p < .05$). This is Garner interference; performance was slower when the irrelevant dimension varied across trials than when it was fixed. These results are shown in Figure 2, which presents performance data that were additionally analyzed in terms of whether the value of the irrelevant dimension repeated (circles) or changed (squares) between trials.

A three-factor analysis of variance (ANOVA) was conducted on filtering-task results, analyzing RTs as a function of range condition (e.g., narrow through wide), relevant-dimension sequence (e.g., repeat or change), and irrelevant-dimension sequence (e.g., repeat or change). There was an effect of the range of pitch variation [$F(2,33) = 8.61, p < .01$], with responses taking longer when pitch variation was greater. There were no reliable effects of whether or not the relevant dimension changed between trials, but there was a marginal effect of whether or not the irrelevant dimension changed [$F(1,33) = 3.66, p < .10$]. However, the interaction between change on the relevant dimension and change on the irrelevant dimension was highly reliable [$F(1,33) = 35.46, p < .01$]. As the range of variation on the irrelevant dimension increased, trials in which the irrelevant dimension changed were increasingly slowed relative to trials in which it repeated.

Figure 3 shows these data classified in terms of stimulus sequence. For these and all other sequence effect data reported in this paper, only correct trials following a correct trial are included in the analyses. The abscissa denotes

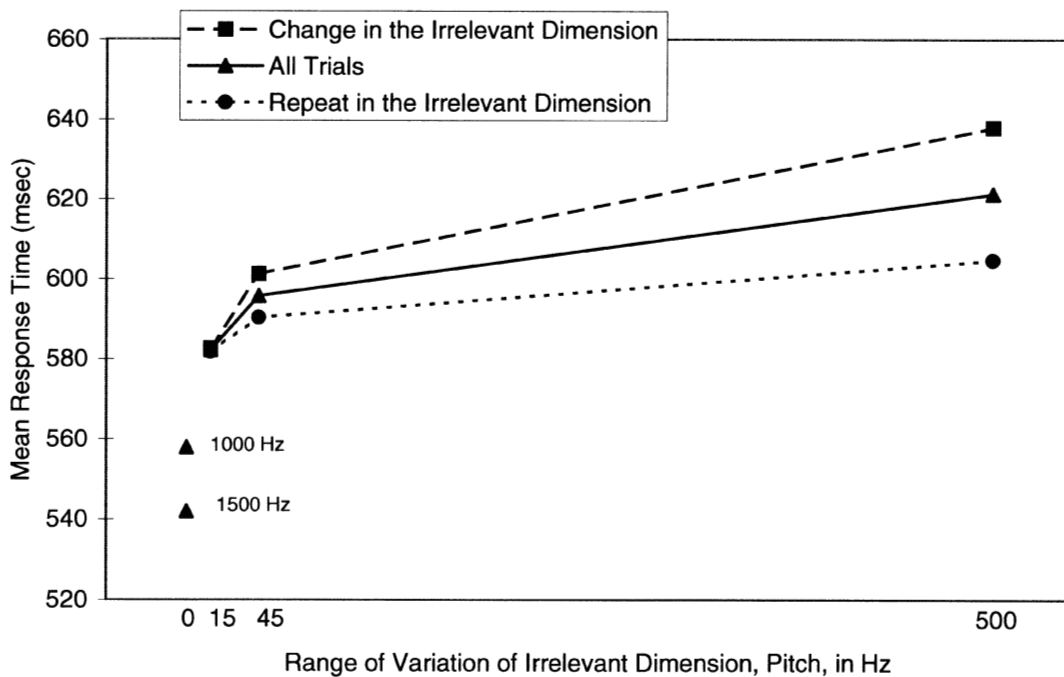


Figure 2. Range effects in Experiment 1A. Mean response times to judge loudness are presented as a function of both the range of variation in pitch (x-axis) and the trial-to-trial sequence of the irrelevant dimension (repeat = circles, change = squares, all trials = triangles). The two separated data points represent response times in the univariate conditions.

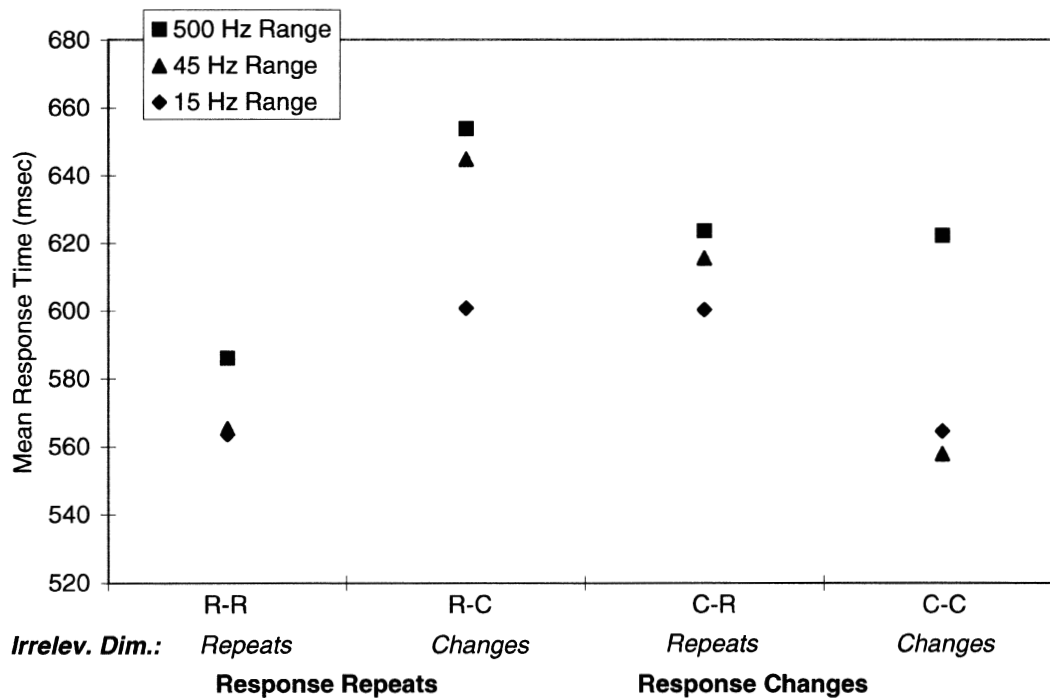


Figure 3. Sequence effects in Experiment 1A. Along the x-axis is stimulus sequence condition (see text), with range condition indicated by the symbols in each column.

the four possible two-stimulus sequence types, from left to right abbreviated R-R, R-C, C-R, and C-C.

Figure 3 shows that stimulus sequence dramatically affects RT. A single pattern characterizes these data for each range condition and for each data set reported ahead in this paper. From fastest to slowest, the conditions were ordered R-R, C-C, C-R, and R-C. The RT for repetition trials, R-R, was fastest of all conditions; however, when the stimulus was preceded by a different stimulus that received the same response, R-C, this response repetition condition was the slowest of all conditions. Responses in the C-C condition, when everything changed, were about the same as those in the R-R condition, when nothing changed. Planned comparisons revealed that both R-R and C-C trials were significantly faster than R-C trials ($ps < .05$), and R-R trials were marginally faster than C-R trials ($p < .10$).

There also was an overall range effect. This is indicated by the tendency for longer RTs in wider range conditions for each sequence. This may reflect an effect of range in addition to the effects of sequence, or there might be sequence effects that we have not measured. The data for each subject are consistent with these summaries across subjects. Other than the finding that some subjects were consistently faster in responding than others, we did not detect any individual differences.

Experiment 1B: Pitch Judgments

Errors. Error rates and RTs for the experimental conditions of Experiment 1B are presented in Table 1. The

overall error rate was 5%. There was no significant correlation between RT and error rate ($r = .58, p > .05$). Error rate was not analyzed as a function of stimulus sequence because there were relatively few errors in each sequence condition (2–4 errors in each, on average).

Response times. RTs are shown in Figure 4. According to planned t tests, performance in the univariate task was faster than that in any other condition (for all tests, $p < .05$), and mean RT in the narrow (4-dB range) condition was not reliably different from that in the medium-narrow (8-dB range) condition. A three-factor ANOVA was conducted, analyzing RT as a function of changes in range of loudness variation, relevant-dimension sequence, and irrelevant-dimension sequence. As in Experiment 1A, RTs increased with increasing range of variation on the irrelevant dimension [$F(3,45) = 8.65, p < .01$]. Overall, trials in which the irrelevant-dimension value repeated from the previous trial were faster than trials in which the irrelevant-dimension value changed [$F(1,45) = 15.16, p < .01$].

The sequence effects for pitch judgments are presented in Figure 5 (labeled in the same way as Figure 3). There was a response-repetition disadvantage, such that trials in which the subjects' response changed from the previous trial were responded to 15 msec faster than trials in which the response repeated [$F(1,45) = 5.20, p < .05$]. As when loudness was judged, there was a significant interaction between change on the relevant dimension (pitch) and change on the irrelevant dimension [$F(1,45) = 34.70, p < .01$]. The pattern of sequence effects observed in Exper-

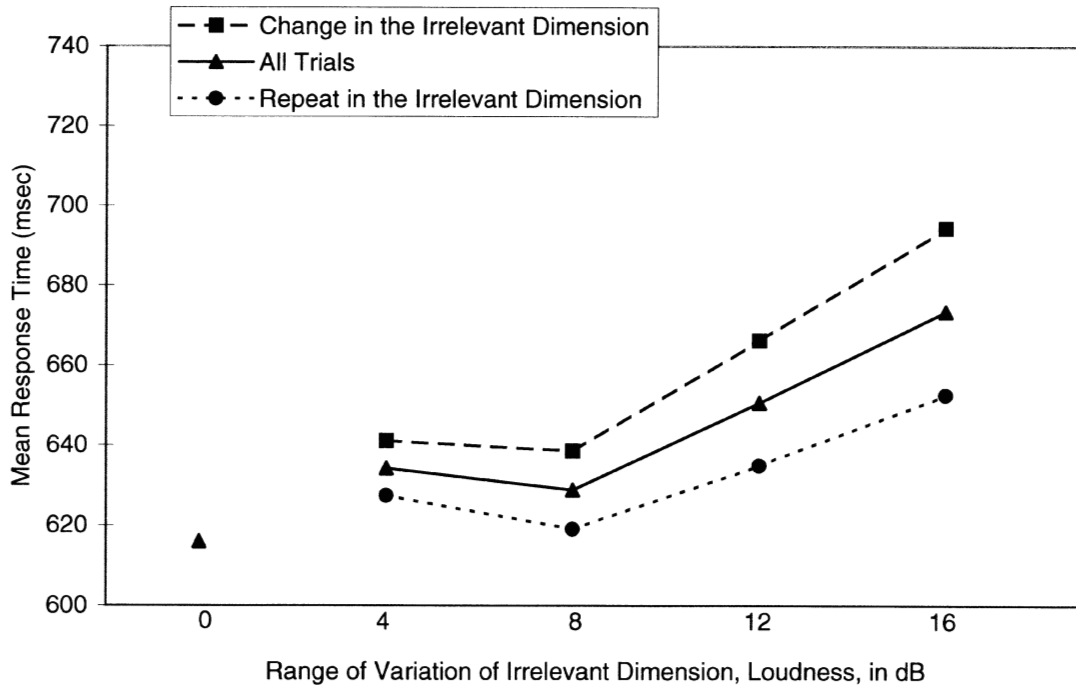


Figure 4. Range effects in Experiment 1B. Mean response times to judge pitch are presented as a function of both the range of variation in loudness (x-axis) and the trial-to-trial sequence of the irrelevant dimension (repeat = circles, change = squares, all trials = triangles). The separated data point represents response time in the univariate condition.

iment 1A was present in these data as well. Responses were fast when neither the stimulus nor the response changed and when both changed. Planned comparisons showed that the R–C condition was the slowest of all conditions (for all pairwise comparisons, $p < .05$), whereas the C–R condition was marginally slower than the R–R condition ($p < .10$).

Range had the same general effect in all conditions here as in Experiment 1A. RTs were slower in general for wider ranges, and the difference in RTs associated with range was numerically greater when loudness changed than when loudness repeated. Post hoc tests revealed this latter difference to be marginally significant [$F(1,3) = 8.17, p < .10$]; this analysis tested the interaction between range condition (narrow or wide) and irrelevant-dimension sequence (repeat or change).

As in Experiment 1A, we were unable to detect any difference in performance across subjects except that some responded more quickly than did others.

Discussion

For judgments of both loudness and pitch, one manipulation—changing the range of variation on an irrelevant stimulus attribute—affects two measures: overall RT and RT as a function of stimulus sequence. These measures are not independent: When the overall range is large, the average difference between successive stimuli is also large.

Range

The averaged RTs increased with increased range of the irrelevant dimension when the subjects classified stimuli on the relevant attribute. These results are consistent with Lockhead (1992b) and Melara and Mounts (1994) and allow extending the conception of orthogonal task interference to include the observation that a greater range of irrelevant variation is associated with more interference or more of whatever else might be the source of the poorer classification performance.

Sequence

Sequence effects are also important. On trials in which both dimensions changed and the response changed (C–C), responses were about as quick (only 10 msec slower on average) as responses to stimulus repetitions when neither dimension changed (R–R) and the response repeated. When only the relevant dimension changed, calling for a different response (C–R), performance was slower than in the C–C and R–R conditions. Slowest of all were trials in which only the irrelevant dimension changed and, thus, the response repeated (R–C). These longest RTs (64 msec slower than R–R) occurred when the only change between trials was not relevant to the task. Together, the results show that neither response repetition nor stimulus attribute repetition accounts for the speed of response. This pattern is consistent with findings reported by Crowder (1989) in a *same–different* task, but

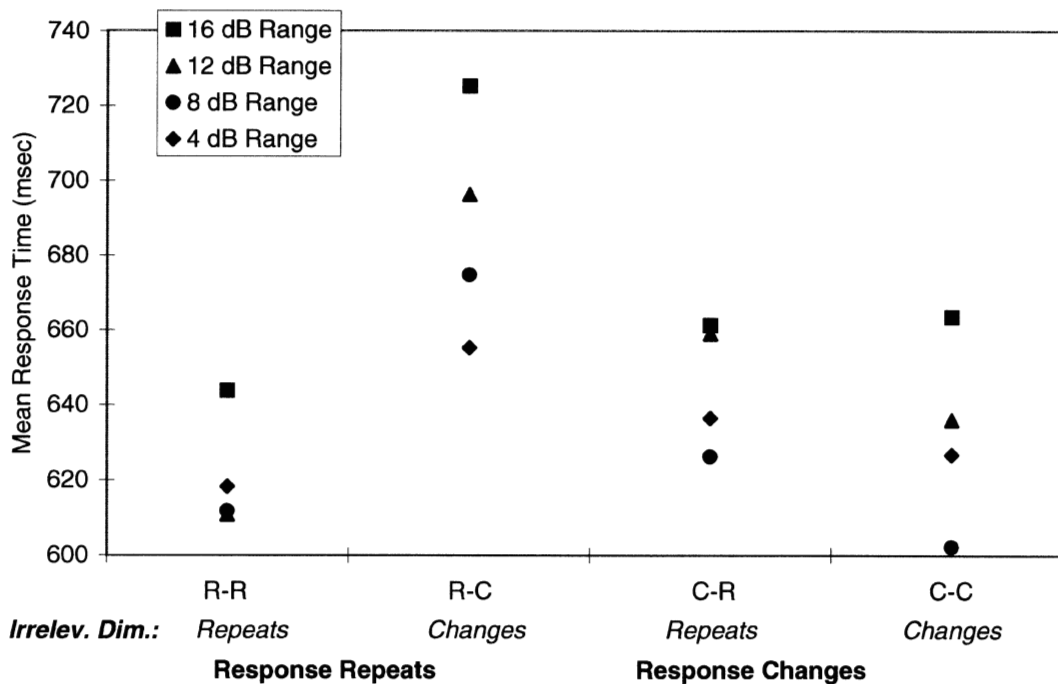


Figure 5. Sequence effects in Experiment 1B. Along the x-axis is stimulus sequence condition (see text), with range condition indicated by the symbols in each column.

it differs from results reported by Felfoldy (1974) and Fletcher (1981) in classification tasks; this discrepancy is considered further in the General Discussion section.

Range and Sequence

Although the pattern of sequence effects was consistent across the ranges tested, the magnitude of these effects does not account for all of the range effect. RTs to every sequence increased with increasing across-tasks range; if sequence fully accounts for range effects, this should not occur. R–R conditions were identical across range conditions, and so RTs should have also been equal across ranges for these sequences. Simply, when the stimulus repeated, RTs to that sequence should have been the same in all range conditions in order for the sequence effects to fully account for the effects of range, and they were not. Thus, Felfoldy's (1974) hypothesis that differences in the sequential structure of tasks accounts for averaged RT differences (i.e., in filtering vs. baseline tasks) cannot be extended to the overall range effect observed in Experiment 1.

EXPERIMENT 2

Within-Task Range Effects in a Filtering Task Having Three Values on the Irrelevant Dimension

The results of Experiment 1 showed a different range effect across filtering tasks having different ranges of the irrelevant dimension. When the range of the irrelevant

dimension was increased, RTs also increased. In Experiment 2, we examined RTs as a function of sequential structure, in order to test whether a similar range effect occurs within a stimulus task as between tasks.

For this purpose, we modified the stimulus set of Experiment 1 by adding a third value on the irrelevant dimension. This increased the number of possible trial-to-trial transitions between stimuli without affecting the dimension to be classified. When the three values of the irrelevant dimension are equally spaced (Experiment 2A), there are three irrelevant transitions: the irrelevant level repeats, changes by ± 1 pitch step, and changes by ± 2 steps. When the values of the irrelevant dimension are unequally spaced, as in Experiment 2B, then there are four irrelevant transitions: 0, 1, 2, and 3 pitch steps.

Because of the pattern of sequence effects in Experiment 1, we anticipated that one of two results would occur. One possibility is that subjects would attend to the presence or absence of change but not to the magnitude of change, resulting in Garner interference but in no range effects. This could be described by Fletcher and Rabbitt's (1978) proposal that subjects may use a "bypass rule": When a stimulus repeats from the previous trial, repeat the response; when a stimulus changes from the previous trial, change the response. This rule allows the subject to "sometimes bypass making a detailed analysis of the stimulus by first comparing it with the preceding stimulus" (Krueger & Shapiro, 1981, p. 242).

This argument was proposed for *same-different* tasks using discrete stimulus values, rather than for classifica-

Table 2
Response Times (RTs; in Milliseconds) and Error Rates
(% Errors) for Sequence Conditions in Experiments 2A and 2B

Sequence	Experiment 2A		Experiment 2B	
	RT	% Errors	RT	% Errors
R-0	433	4.0	502	7.6
R-1	497	6.3	569	11.0
R-2	498	7.8	572	16.0
R-3			601	12.4
C-0	464	9.8	544	10.0
C-1	461	5.3	532	7.6
C-2	431	5.0	520	7.4
C-3			508	10.2

Note—For sequences, the letter indicates whether the relevant dimension repeated (R) or changed (C), and the digit indicates the magnitude of irrelevant-dimension change.

tion tasks using continuous dimensions, which makes the rule difficult to apply here. However, a bypass rule for classification task would have the following structure: When subjects do not detect change from the previous trial, they tend to repeat the previous response; when they do detect change from the previous trial, they change the response. So, trials on which the response repeats because the relevant level repeats but the stimulus changes because the irrelevant level changes would be slow, relative to all other trials. This account makes no predictions regarding range effects.

Another possible result is that range effects will be present in the data. This would occur if the interaction observed in Experiment 1 increases in strength with increasing magnitudes of irrelevant change. Responses would be slowed (following the suggestion of Krueger & Shapiro, 1981) when the above bypass rule was incorrect, as when stimulus change was detected but the response repeats or when no change was detected but the response changes. Increasing magnitude of irrelevant-dimension change would increasingly slow response repetitions and speed response changes.

A third possibility, based on Dixon and Just's (1978) result in *same-different* tasks, is that responses are slowed more by larger changes in the irrelevant dimension, but only when the relevant dimension level repeats. Their interpretation is that subjects mentally normalize stimuli when they differ on the irrelevant dimension.

Experiments 2A and 2B used more than two stimulus levels on the irrelevant dimension but used only two levels on the relevant dimension. This means that the relevant dimension (loudness) could only repeat (R) or change (C) between trials, whereas the irrelevant dimension could change by different magnitudes. Figure 1B shows a stimulus set similar in form to that used in Experiment 2B, with three unequally spaced values on the irrelevant dimension. If a stimulus repeated from the previous trial, that sequence would be expressed as the *R-0* category, in that the relevant dimension repeated and the irrelevant dimension changed by zero steps (i.e., it repeated). If Stimulus B from Figure 1B was presented and

followed by Stimulus A, the sequence would be *R-1*, since the relevant level repeated and there was a one-step change on the irrelevant dimension. Likewise, a change from Stimulus A to Stimulus F would be *C-3*. This categorization only addressed the magnitude of change on the irrelevant dimension, because the direction of change was not found to matter in regard to RTs or errors.

Method

Subjects

Nine subjects (4 in Experiment 2A, and 5 in Experiment 2B) from the Duke University community participated in Experiment 2. All subjects were paid \$10. Each subject participated for 1 h.

Stimuli

Stimuli were generated by a SoundBlaster-compatible soundcard and were heard over loudspeakers. The tones were created using the Experimental Run Time System (ERTS; Berisoft Corp.). The tones were 100-msec sine waves with ramped onset and offset.

In Experiment 2A, there were two intensity levels (judged), 76 and 80 dB, and three levels of pitch (not judged), 523, 554, and 587 Hz. The pitch levels correspond to MIDI notes (e.g., the musical notes Middle C, C#, and D). This stimulus set provided two sequences on the relevant dimension, repeat and change, and three sequences on the irrelevant dimension, repeat, change by 1 step (change to an adjacent frequency value), and change by 2 steps (change from the highest frequency to the lowest, or vice versa).

In Experiment 2B, stimuli were identical to those used in Experiment 2A, except that the highest frequency value was changed from 587 Hz to 623 Hz (e.g., D#). This set again provided two possible sequences on the relevant dimension: repeat and change. It also provided four possible magnitudes of change along the irrelevant dimension. The stimulus could repeat along the irrelevant dimension, or the irrelevant dimension could change by 1 MIDI step (between the lowest two values), by 2 steps (between the highest two values), or by 3 steps (between the lowest value and highest values). The total range of irrelevant variation increased from 64 Hz in Experiment 2A to 100 Hz in Experiment 2B.

Task

The subjects classified the tones on the basis of loudness and were instructed to ignore random variation in pitch between trials. To indicate a response, the subjects pressed a key on a custom-made response box designed to work with ERTS, which allows better than millisecond precision of stimulus presentation and response measurement. Response keys were counterbalanced across subjects.

Procedure

All subjects were tested in a quiet, isolated room. Instructions advised them to classify the stimuli on the basis of loudness and to respond as quickly as possible while maintaining a low error rate.

The subjects first received a practice block of 102 trials. Each practice trial consisted of a stimulus for 100 msec, a response interval of 900 msec, feedback displayed for 500 msec, and a 1,000-msec intertrial interval. The subjects were provided one of three types of feedback: "correct," "incorrect," and "respond more quickly" (if the response was not within the designated interval). Practice trial data were not analyzed. The subjects then participated in six test blocks. Each test block consisted of 24 practice trials followed by 102 test trials. The test trials were identical to the practice trials, except for removal of the feedback interval and information. That is, they consisted of the 100-msec tone, the 900-msec response interval, and the 1,000-msec intertrial interval. Stimuli were presented randomly and equally often.

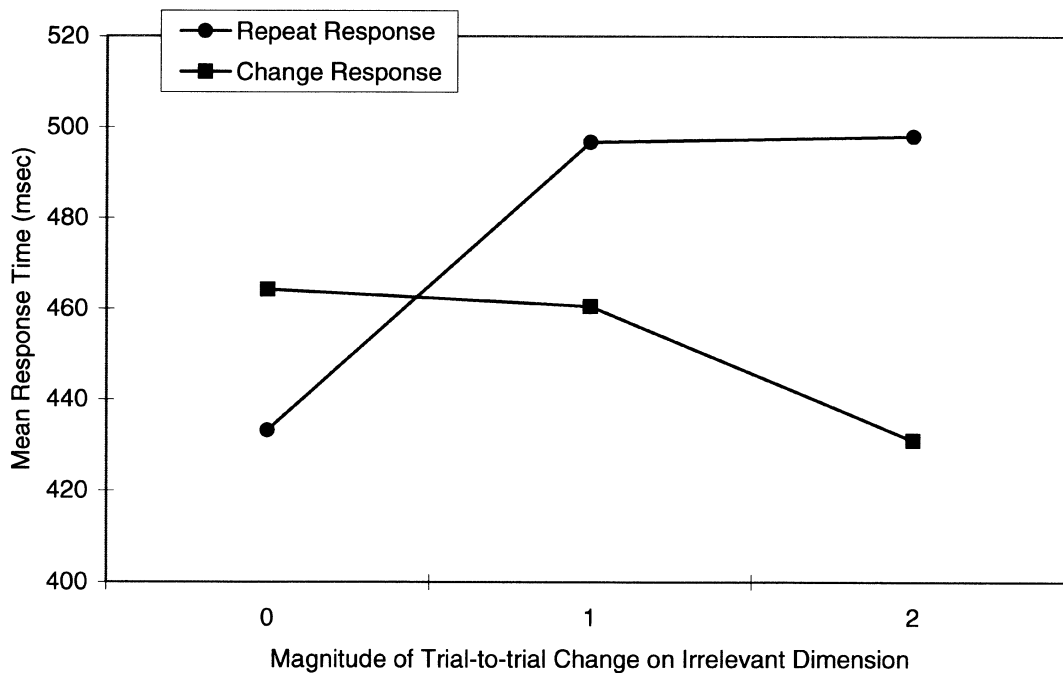


Figure 6. Sequence effects in Experiment 2A. Mean response time is a function of trial-to-trial change on both the irrelevant dimension (x-axis) and the relevant dimension (repeat = circles, change = squares).

Results

Experiment 2A

Errors. Error rates and RTs for Experiment 2A are presented in Table 2. Error rates averaged 6%. Errors within sequence conditions did not trade off with RT ($r = .54, p > .05$).

Response times. For Experiments 2A and 2B, two-factor repeated measures ANOVAs were conducted. Only correct responses that followed a correct response were included in the analyses. RTs were analyzed as a function of relevant-dimension sequence (repeat or change) and irrelevant-dimension sequence (magnitude of irrelevant change). There was no main effect of trial-to-trial change on the irrelevant dimension [$F(2,6) = 3.53, p > .05$]; overall, responses on trials in which the irrelevant dimension changed were about as fast as trials in which the irrelevant dimension repeated. Response-repetition trials were not significantly faster than response-change trials [$F(1,3) = 1.76, p > .05$].

Figure 6 shows an interaction in RTs between when there was change on the relevant dimension and when there was change on the irrelevant dimension [$F(2,6) = 13.38, p < .01$]. Responses were faster to stimulus repetitions (R-0) on trials in which the irrelevant dimension did not change than on trials in which the relevant dimension (and the response) changed (C-0). This ordering reversed on trials with change on the irrelevant dimension, such that response repetitions took longer than response

changes. This interaction replicated the results of Experiment 1.

The magnitude of this sequence effect increased numerically with increasing changes on the irrelevant dimension. For zero-step changes (repetitions of the irrelevant level), "same" responses were 31 msec faster than "change" responses. The reverse occurred when there was a one-step change on the irrelevant dimension, when "change responses" were faster than "same" responses by 36 msec. When the irrelevant dimension changed by two steps, "change" responses were 67 msec faster than same responses.

A second ANOVA analyzed RTs of only trials in which the irrelevant dimension changed from the previous trial (i.e., the ANOVA excluded R-0 and C-0 trials). For these trials, there was a marginally significant effect of response repetition [$F(1,3) = 6.79, p < .10$] but no significant effect of the magnitude of irrelevant-dimension change [$F(1,3) = 1.06, p > .05$] and no significant interaction between the factors [$F(1,3) = 1.85, p > .05$]. This analysis indicates, for this stimulus set, that the observed within-conditions range effect was not present when only irrelevant-change trials were analyzed.

Experiment 2B

Errors. Error rates and RTs are presented in Table 2. Error rates averaged 10%. There was a positive correlation between RTs and error rates ($r = .72, p < .05$).

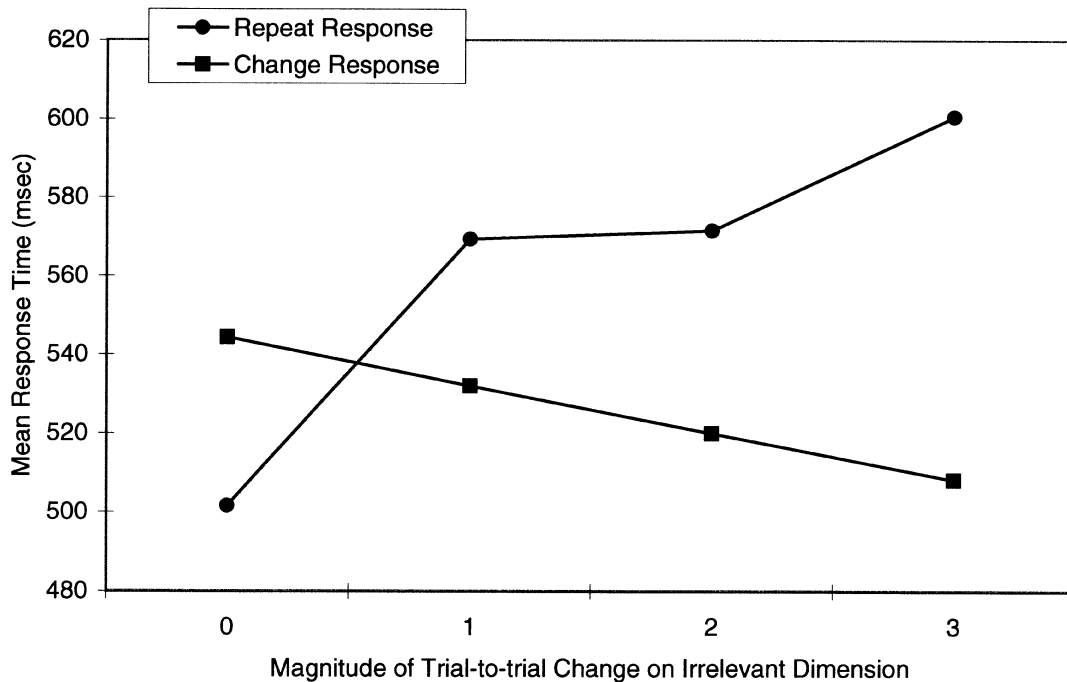


Figure 7. Sequence effects in Experiment 2B. Mean response time is a function of trial-to-trial change on both the irrelevant dimension (x-axis) and the relevant dimension (repeat = circles, change = squares).

Response times. Figure 7 illustrates the findings from Experiment 2B, with RT plotted both as a function of type of change on the relevant dimension and of magnitude of trial-to-trial change on the irrelevant dimension (x-axis). For these analyses, only correct responses that followed a correct response were included in the results.

There was an effect of change on the irrelevant dimension, such that responses were slower when the irrelevant-dimension change was large than when the irrelevant dimension repeated [$F(3,12) = 4.96, p < .05$]. There was an effect of change on the relevant dimension, such that response repetitions were 26 msec slower, on average, than response changes [$F(1,4) = 8.29, p < .05$]. These main effects interacted [$F(3,12) = 27.85, p < .01$].

The interaction observed previously was manifested in these data as well: Increasing magnitude of change on the irrelevant dimension interfered with responding on trials when the relevant dimension repeated, but it facilitated responding on trials when the relevant dimension changed. The size of this interaction ranged from a 42-msec repetition advantage when there was no irrelevant change to a 93-msec repetition disadvantage for the largest trial-to-trial irrelevant changes, for a total estimated effect of 135 msec.

The interaction was tested in only trials in which the irrelevant dimension changed from the previous trial (i.e., through an ANOVA excluding R-0 and C-0 trials). This interaction was significant [$F(2,8) = 13.80, p < .01$], in-

dicating a within-conditions range effect even when no-change trials were excluded from the analysis.

Discussion

For Experiments 2A and 2B, RT to a given stimulus was greatly affected by trial-to-trial changes. Also, the interaction found in Experiment 1 again was present in these experiments. Change on the irrelevant dimension slowed responses when the relevant dimension repeated, but it facilitated responses when the relevant dimension changed.

Additionally, there was a within-conditions range effect in Experiment 2B: As the magnitude of irrelevant-dimension change increased, RT to response repetitions slowed, relative to response changes. This within-conditions range effect was in the same direction but was not significant in Experiment 2A, which had a somewhat smaller overall range of variation (2 steps vs. 3 steps) and fewer different irrelevant-dimension sequences (3 transitions vs. 4 transitions) than Experiment 2B.

These results for classification tasks, in which subjects are instructed to sort according to a feature, have a similar pattern to many results in *same-different* tasks, where subjects are instructed to compare stimulus objects (see Farrell, 1985, for a review). Subjects may do the same things in these different situations. Whatever the underlying process, the finding of range effects within conditions allows rejecting the original bypass rule as an ac-

count of classification performance, because RT does vary as a function of the magnitude of trial-to-trial irrelevant-dimension change. However, the suggestion by Krueger and Shapiro (1981) is consistent with these results; subjects still might use a strategy similar to the bypass rule, such that these range effects occur because some sequential changes are more readily detected than others. This was tested and rejected in Experiment 3.

EXPERIMENT 3 Within-Task Range Effects With Many Values on the Irrelevant Dimension

Experiment 2 demonstrated range effects within a single orthogonal (filtering) sorting task, revealed an interaction between effects of relevant and irrelevant change (irrelevant trial-to-trial change slows response repetitions but speeds response changes), and showed that this interaction increases in magnitude with increasing amounts of irrelevant change. This extended Experiment 1 by changing from two to three levels of the irrelevant dimension. In both experiments, it is conceivable that each stimulus could have been held individually in memory such that, as a stimulus was presented, the subject identified it directly rather than selectively attending to the relevant dimension. If such an identification strategy was employed, then the noted range effects might have been due to imperfect use of that strategy (e.g., more confusions when the range was larger, as occurs in univariate-classification studies; cf. Gravetter & Lockhead, 1973), rather than due to trial-to-trial changes between attribute levels or stimuli directly.

Additionally, the use of few stimuli might have biased the subjects to use something like the bypass rule of Fletcher and Rabbitt (1978), rather than to directly identify the level of the relevant attribute. In Experiment 1, that rule could provide the correct response on 75% of the trials (stimulus repetitions and response changes). If the subjects did use the rule, they also did something else, since performance was 95% correct. Adding more levels of the irrelevant dimension reduced the effectiveness of that method even further. With 12 levels on the irrelevant dimension, the bypass rule provided the correct response for only 13 of the 24 possible stimulus sequences.

Thus, we expanded the range and number of alternatives on the irrelevant dimension to 12 levels spanning a full octave. These perceptually equal pitch steps provide for more detailed measurement, relative to Experiments 1 and 2, as to how irrelevant-dimension trial-to-trial change relates to classification performance.

Two stimulus sets were used in this Experiment 3, each incorporating many values on the irrelevant dimension. The subjects again classified auditory stimuli on the basis of loudness while ignoring irrelevant pitch variation. The stimulus set in Experiment 3A had 24 stimuli generated by two values of the relevant dimension (loudness), each presented with 12 equally spaced values of the irrelevant

dimension (pitch). The set in Experiment 3B had 16 stimuli; there are two values of the relevant dimension and eight values on the irrelevant dimension, with the highest and lowest of these eight values separated from the other six by a relatively large amount.

We adopted these stimulus sets because the large number of values makes a direct object identification strategy unlikely and provides for testing effects of stimulus spacing on sequence effects. The sets differed in overall form: Items in Experiment 3A were equally spaced, whereas those in Experiment 3B had a central cluster of stimuli with outlying stimuli. However, the sequential transitions between stimuli were similar in both cases; both sets provided for stimulus repetitions, 1-step changes, 11-step changes, and so forth. Comparing the patterns of sequence effects in the two experiments enabled us to measure whether performance depended on set manipulations in classification, as it did in the *same-different* tasks reported by Dixon and Just (1978).

Using univariate stimuli spaced similarly to the spacing of the irrelevant-dimension values on Experiment 3B, Parducci and Perrett (1971) and Gravetter and Lockhead (1973) reported, respectively, that stimulus range and criterial range predict average performance. Using a bivariate stimulus set structured similarly to that of Experiment 3B, Luce and Green (1978) found that performance for judging outlier stimuli was worse than for judging stimuli in the central cluster. All of these results, none of which was presented with sequential measures, might be accounted for by differences in sequence.

Method

Subjects

Twenty subjects (10 in Experiment 3A, and 10 in Experiment 3B) from the Duke University community were paid \$10 to participate in the experiment. Each subject participated for 1 h.

Stimuli

Stimuli were identical in form and presentation to those from Experiment 2, except that there were additional values of the irrelevant dimension, pitch.

In Experiment 3A, there were two intensity levels on the judged dimension of loudness, 76 and 80 dB, and 12 levels on the non-judged dimension of pitch. These pitch levels were the 12 MIDI notes of an octave, with each level representing one semitone from middle C (523 Hz) to B (987 Hz). This stimulus set provided two sequences on the relevant dimension, repeat or change, and 12 sequences on the irrelevant dimension, from a 0-step change (repeat the irrelevant-dimension value) to an 11-step change (from lowest to highest pitch, or highest to lowest).

In Experiment 3B, the stimulus set from Experiment 3A was repeated except that four of the frequency values were removed; with the frequencies ranked from lowest to highest, numbers 2, 3, 10, and 11 were excluded. The resulting set had six frequencies in the middle of the range, plus one outlier at a lower frequency and one outlier at a higher frequency.

Task

The task was identical to that used in Experiment 2: The subjects classified the tones on the basis of loudness and were instructed to ignore random variation in pitch between trials.

Table 3
Response Times (RTs; in Milliseconds) and Error Rates
(% Errors) for Sequence Conditions in Experiments 3A and 3B

Sequence	Experiment 3A		Experiment 3B	
	RT	% Errors	RT	% Errors
Response Repetitions				
R-0	502	6.1	439	6.3
R-1	521	7.7	470	9.1
R-2	555	5.8	492	9.0
R-3	571	7.4	492	9.9
R-4	584	6.8	514	11.2
R-5	591	9.6	517	12.0
R-6	602	10.0	533	7.2
R-7	610	10.4	537	14.0
R-8	638	8.1	546	15.6
R-9	612	11.7		
R-10	658	16.3		
R-11	645	11.9	604	17.6
Response Changes				
C-0	588	13.1	524	18.7
C-1	560	8.4	486	11.6
C-2	560	10.4	481	10.4
C-3	568	8.6	490	11.9
C-4	568	8.3	493	8.1
C-5	550	8.4	498	11.1
C-6	563	8.5	501	11.7
C-7	564	8.2	481	10.2
C-8	569	10.9	495	9.2
C-9	571	11.9		
C-10	528	7.2		
C-11	579	7.0	458	18.2

Note—For sequences, the letter indicates whether the relevant dimension repeated (R) or changed (C), and the digit indicates the magnitude of irrelevant-dimension change.

Procedure

The procedure was similar to that used in Experiment 2. The only difference was in the number of trials per testing block: So that every stimulus was presented a similar number of times, the number of trials had to be an integer multiple of the number of stimuli. For Experiment 3A, the practice block consisted of 72 practice trials, whereas each testing block consisted of 24 practice trials and 96 test trials. For Experiment 3B, the practice block contained 64 practice trials, and each testing block contained 16 practice trials and 96 test trials. Stimuli were presented randomly and equally often. Therefore, successive stimulus differences were not uniform: There were more small trial-to-trial changes in pitch than large trial-to-trial changes in pitch.

Results

Experiment 3A

Errors. Error rate and RT information for Experiment 3A are presented in Table 3. Error rates in Experiment 3A averaged 9%. There was no significant correlation between error rates and RTs across subjects ($r = .16$, $p > .05$), indicating that RT differences among subjects were not attributable to speed-accuracy tradeoffs. We also analyzed error rates as a function of range condition. For trials in which the relevant dimension, loudness, repeated (e.g., R-0, R-11), error rates positively correlated with both RTs ($r = .77$, $p < .01$) and the magnitude of trial-to-trial change (i.e., the number of unit steps of

change; $r = .83$, $p < .01$). Error rates generally increased with increasing trial-to-trial change when the relevant dimension repeated.

For trials in which the relevant dimension changed (e.g., C-0, C-11), error rates were not significantly correlated with RTs ($r = .53$, $p > .05$) or with the amount of irrelevant-dimension change ($r = -.37$, $p > .05$). However, the RTs for these sequence conditions did not significantly vary (see below for analysis), so the reduced correlation between error rates and RTs might have been attributable to the reduction in RT variance.

Response time. Figure 8 presents RT as a function of trial-to-trial stimulus sequence. The x -axis indicates the magnitude of change between trials on the irrelevant dimension, pitch. Zero on the abscissa represents trials in which the irrelevant dimension did not change from the previous trial; similarly, 11 indicates trials in which the irrelevant dimension changed by the largest amount possible.

A two-factor repeated measures ANOVA, with factors of relevant-dimension sequence and irrelevant-dimension sequence, showed no main effect of relevant-dimension sequence; response repetitions were about as fast as response changes [$F(1,9) = 1.13$, $p > .05$]. There was a main effect of irrelevant-dimension sequence, with responses increasingly slowed by increasing magnitude of change on the irrelevant dimension [$F(11,99) = 2.98$, $p < .01$]. However, a highly significant interaction [$F(11,99) = 6.64$, $p < .01$] superseded interpretation of any main effect; this was also observed in Experiments 1 and 2. When the level of the relevant dimension repeated, RTs were slowed more by larger increases in the magnitude of irrelevant change. This effect was essentially linear with the logarithmic frequency of the x -axis ($r = .96$, $p < .01$; slope = 12.7 msec/step, intercept = 521 msec). However, when the relevant dimension changed, RTs were unaffected by the magnitude of change on the irrelevant dimension ($r = -.26$, $p > .05$; slope = -1.1 msec/step, intercept = 569 msec).

For trials in which the relevant dimension changed, a post hoc test compared RTs when the irrelevant dimension did not change (0-step) to when it did change (1- to 11-step). There was only a marginal effect [$F(1,9) = 3.59$, $p < .10$], further indicating that irrelevant change had little effect when the relevant dimension changed.

The interaction above was also tested in only those trials in which the irrelevant dimension changed from the previous trial (i.e., through ANOVA excluding R-0 and C-0 trials). This interaction was also significant [$F(10,99) = 4.32$, $p < .01$]. Thus, the within-conditions range effect was significant even when no-change trials were excluded from the analysis.

The observed interaction was not only significant, it was of large magnitude. When there was no change on the irrelevant dimension, relevant-dimension repetitions were 85 msec faster than relevant-dimension changes. But, for large changes on the irrelevant dimension, relevant-

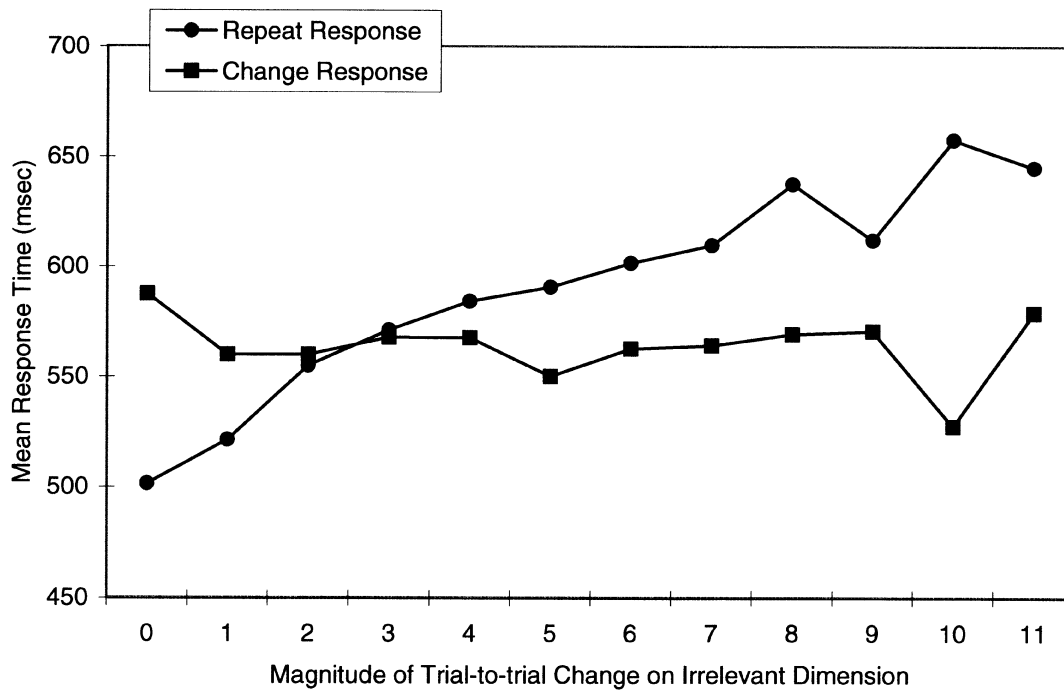


Figure 8. Sequence effects in Experiment 3A. Mean response time is a function of trial-to-trial change on both the irrelevant dimension (x-axis) and the relevant dimension (repeat = circles, change = squares).

dimension changes were faster by an approximately similar margin. This reversal was nearly entirely driven by the effect on relevant-dimension repetitions, which showed a range effect of about 150 msec across the irrelevant-change sequences tested.

Experiment 3B

Errors. Error rate and RT information for Experiment 3B are presented in Table 3. Error rates averaged 12%. Errors correlated positively with RTs across subjects ($r = .60, p = .05$).

When the relevant dimension repeated, error rates were positively correlated with both RTs ($r = .83, p < .01$) and the magnitude of trial-to-trial change (i.e., the number of unit steps of change; $r = .85, p < .01$).

When the relevant dimension changed, there was no correlation between error rates and RTs ($r = .04, p > .05$), nor was there any correlation with the amount of irrelevant-dimension change ($r = -.42, p > .05$). As in Experiment 3A, this lack of a correlation between error rates and RTs might have been attributable to the small variance in the RT data (see below), which were essentially constant across the range of variation.

Response times. The RTs are summarized in Figure 9. The overall structure of the data was similar to that from Experiment 3A. An ANOVA similar to that from Experiment 3A was conducted, but the 11-step changes were excluded from only this analysis due to the small number of trials per subject, relative to the other conditions. There was no main effect of change on the relevant

dimension [$F(1,9) = 0.98, p > .05$], but there was a main effect of change on the irrelevant dimension [$F(8,72) = 3.17, p < .01$] and there was a significant interaction between change on the irrelevant dimension and change on the relevant dimension [$F(8,72) = 8.34, p < .01$]. As the magnitude of trial-to-trial change on the irrelevant dimension increased, RTs to trials in which the relevant dimension repeated increased linearly (given the logarithmic frequency scale on the x-axis; $r = .98, p < .01$; slope = 13.1 msec/step, intercept = 453 msec). RTs had a slight negative trend with increasing irrelevant change when the relevant dimension also changed ($r = -.60, p < .05$; slope = -3.0 msec/step, intercept = 505 msec). This latter correlation was largely driven by the C-11 trials (those with the largest trial-to-trial change but the fewest occurrences per subject). With those trials excluded from this analysis, the correlation became nonsignificant ($r = -.29, p > .05$; slope = -1.4 msec/step, intercept = 500 msec).

A similar post hoc test to that from Experiment 3A, testing whether RT to trials in which the relevant-dimension changed was different when the irrelevant dimension repeated from when the irrelevant dimension changed, was significant [$F(1,6) = 6.41, p < .05$]. Finally, the original interaction was again significant when only those trials in which the irrelevant dimension changed from the previous trial (i.e., through ANOVA excluding R-0 and C-0 trials) were tested [$F(7,63) = 2.88, p < .05$].

As for Experiment 3A, the magnitude of this interaction was notably large. Here, an 85-msec response-repetition

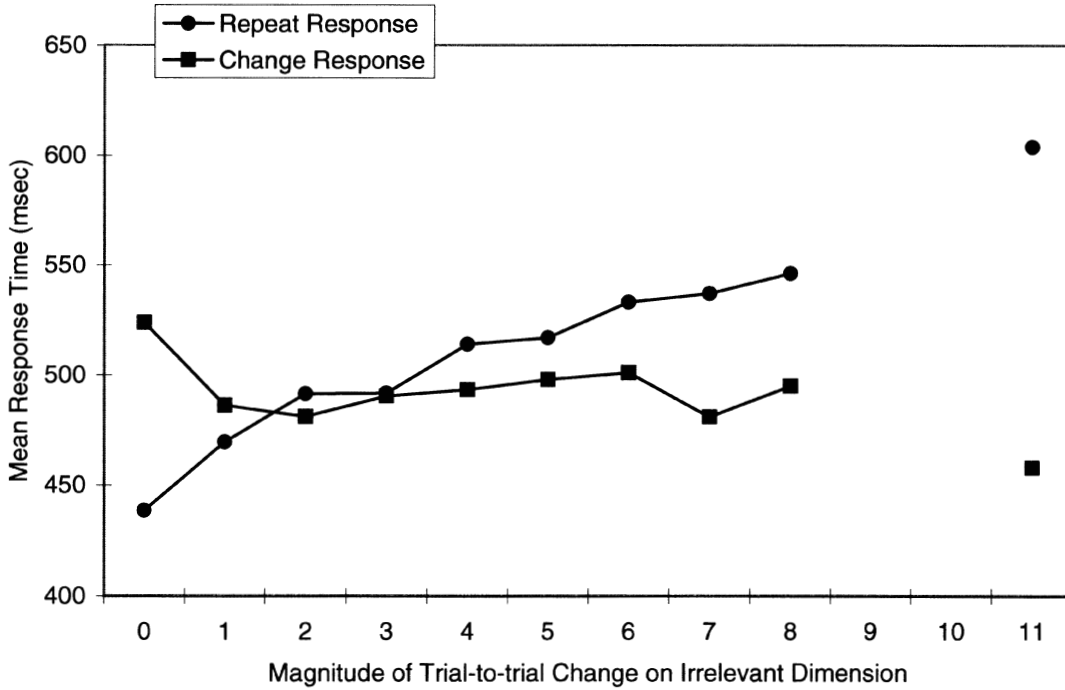


Figure 9. Sequence effects in Experiment 3B. Mean response time is a function of trial-to-trial change on both the irrelevant dimension (x-axis) and the relevant dimension (repeat = circles, change = squares).

advantage with no irrelevant change shifted to about a 55-msec disadvantage at 7- and 8-step changes, for a 140-msec difference. For the 11-step change, this difference was 150 msec, but this might be evaluated cautiously because it was based on few trials; these sequences occurred only when the highest pitched stimulus was followed by the lowest pitched stimulus, or the reverse.

We separately analyzed the data as a function of whether the stimulus was an outlier on the irrelevant dimension (i.e., it had either the highest or the lowest value) or was in the middle cluster (i.e., it had any of the central six values). These data generally fit linear functions similar to that of the overall data from Experiments 3A and 3B. When the relevant-dimension level repeated from the previous trial, responses were similar for both types of stimuli: For outlier stimuli, $r = .92$ ($p < .01$), with slope of 12.2 msec/step and an intercept of 455 msec; for cluster stimuli, $r = .85$ ($p < .01$), with slope of 10.6 msec/step and an intercept of 462 msec.

On trials in which the relevant dimension changed, the patterns were also similar for both types of stimuli. There was a slight negative slope for cluster stimuli ($r = -.55$, $p < .05$; slope = -2.9 msec/step, intercept = 505 msec), with the irrelevant-dimension repetitions slower than any other sequence condition. There was a larger negative slope for outlier stimuli ($r = -.73$, $p < .01$; slope = -7.0 msec/step, intercept = 529 msec), which might have been due more to a greater slowing of irrelevant-

dimension repetitions (R=0 trials) than to any systematic effects across the other irrelevant transitions.

Discussion

In Experiments 3A and 3B, the range effect of variation in the irrelevant dimension on RT was reliable and essentially linear (adjusted for the log scale of frequency on the x-axis). For these particular data, when the level of the relevant dimension repeated (and thus the response repeated), each additional unit of irrelevant change resulted in a similar increase in the average RT, approximately 13 msec. The irrelevant dimension varied over 11 semitone steps, a 1-octave range. It cannot be known whether this 13-msec slope was associated with number of steps, size of steps, range, or some combination of these, and it cannot be known whether the linear effect found here holds generally. But it is clear that the time that subjects take to repeat a response is longer when successive stimuli differ more on the nominally irrelevant dimension.

Variation between trials in the level of the relevant dimension also matters to performance. When the relevant level changed (and thus the response changed), RTs to judge loudness were essentially unaffected by any change in pitch, the irrelevant dimension. The RTs were similar for all magnitudes of irrelevant-dimension change, with one exception. For C=0 trials (the relevant level changed and the irrelevant level repeated), responses tended to be slower than for other response-change trials; this was

statistically significant in Experiment 3B and marginally significant in Experiment 3A according to post hoc testing.

We emphasize two points here: First, the magnitude of irrelevant change between trials directly influenced responses, but only on trials in which the response repeated (and thus when the relevant-dimension level repeated). Irrelevant change did not affect performance when the response changed (and thus when the relevant-dimension level changed), except for a suggestion of slowing on C-0 trials.

Second, performance when the relevant dimension changed (change the response) was about the same as that when the relevant dimension repeated (repeat the response). Thus, it appears that whether or not information from the irrelevant dimension affects performance depends on the characteristics of the relevant dimension.

The asymmetry in the effects of irrelevant-dimension change found in Experiment 2 was again present in Experiment 3. Although irrelevant-dimension change slowed response repetitions, it did not speed response changes. The only difference between RTs among response-change situations was the suggestion of a slowing of C-0 trials relative to other trials.

While these results have not been reported previously for classification tasks, our classification results are quite consistent with results reported by Dixon and Just (1978) for *same-different* tasks. The findings are not consistent with the suggestion of Krueger and Shapiro (1981) that modifies the bypass rule of Fletcher and Rabbitt (1978). According to this suggestion, response repetitions should be slowed by increasing irrelevant-dimension change, which was found, but response changes should be speeded by increasing irrelevant-dimension change, which was not found. The results also cannot be fully explained by Dixon and Just's (1978, pp. 37-38) suggestion that "subjects mentally equate the two stimuli on the relevant dimension before deciding that they are identical along the irrelevant dimension." This account, called *normalization*, does not address why performance is not affected by irrelevant range when consecutive stimuli differ on the relevant dimension, when one might again expect normalization to be needed.

A process that might account for all of the results is the following: We suppose that subjects hold the most recent stimulus in memory and compare that memory with the perception of the next stimulus. If this comparison indicates that the stimuli are identical, then the previous response is repeated (this has analogy to the "response-selection shortcut" of Pashler & Baylis, 1991), so response repetitions should be relatively fast, as they are. When this comparison indicates that the present stimulus is different from the previous one, then the subject prepares a different response from that on the previous trial.

For both stimulus-repetition trials and response-change trials, such a process predicts relatively fast, accurate responding, and this is what we found in Experiment 3. On

trials in which the relevant level repeats but the irrelevant level changes, then this newly prepared "different" response (due to the change in the stimulus from the last trial) must be halted, and the repeat response must be reinstated (cf. St. James & Eriksen, 1991), and this takes time. Furthermore, this reinstatement requires more time when successive stimuli are more different, perhaps because attention has been moved farther from the memory stimulus by the very different current stimulus (cf. Lockhead, 1992a). This, too, is what we found.

The two stimulus sets in Experiment 3 differed in structure: The members of one set were equally spaced throughout the range of irrelevant values, whereas those of the other set had a central cluster of stimuli and some extreme members. However, the two patterns of results were highly similar. We conclude that the subjects processed the individual stimuli and stimulus differences in the same way in both sets, irrespective of this difference in their compositions. Consistent with this, the processing of outliers within the clustered set (Experiment 3B) was not noticeably different from that of clustered stimuli, once sequence is accounted for. RTs to both outlier and clustered stimuli increased uniformly across increasing amounts of trial-to-trial change in the irrelevant dimension.

These results suggest modifying Luce and Green's (1978) conclusions based on a similar task by showing that stimulus sequence, not relations between stimuli in a total stimulus set, determines classification performance. In a classification task in which stimuli are presented equally often, the average trial-to-trial change is necessarily larger to stimuli with extreme values on a dimension than to stimuli with more central values. The highest pitched stimuli in Experiment 3A, for instance, will follow each possible pitch transition (from 0 to 11 steps) equally often, whereas stimuli with more medial values of pitch will only have smaller pitch transitions (e.g., from 0 to 6 steps). We also know that RTs are longer to larger stimulus transitions than to smaller transitions (King, Gruenewald, & Lockhead, 1978). Therefore, even though RTs as a function of sequence were similar for both outlier and cluster stimuli (as demonstrated in Experiment 3B), the RTs averaged across all trials should be longer for the outlier stimuli than for the central stimuli, because there are many more instances of large differences between successive stimuli when outliers are judged than when central stimuli are judged. Although inconsistent with Luce and Green's (1978) interpretation, this is consistent with the results of their Experiment 2 that show worse performance (i.e., higher error rates) for judging outlier stimuli than for judging cluster stimuli. We extend their work by showing that judgments of stimuli in the clustered set (Experiment 3B) are similar to those in the equally spaced stimulus set (Experiment 3A). We conclude that RT to any stimulus depends on the trial-to-trial transition, not on its identity (e.g., cluster or outlier stimulus).

GENERAL DISCUSSION

There were marked range and sequence effects in bivariate classification tasks, and these were highly similar to range and sequence effects found in univariate classification tasks. In general, the differences between all stimuli in a set of items affected performance in classifying any particular item (set or range effects), and the difference between successive stimuli further affected performance in classifying each stimulus (sequence effects). For both univariate and bivariate stimulus sets, identification and classification performance was poorer when the overall range was larger and when the difference between successive stimuli was larger.

The experiments reported here provide new or supporting evidence for six specific conclusions:

1. We replicated previous findings (Lockhead, 1992b; Melara & Mounds, 1994) of across-tasks range effects of an irrelevant dimension on RT. For all stimulus sets that have been examined and reported in the literature, increasing the range of the irrelevant dimension slows overall responding (cf. Figures 2 and 4).

2. There are sequence effects in two-by-two classification tasks (Experiment 1) and in univariate tasks (Holland & Lockhead, 1968). For orthogonal (filtering) classification tasks, responses are generally faster when both stimulus dimension levels change in value or both repeat in value than when one dimension level changes and the other repeats. For *same-different* tasks, Dixon and Just (1978) reported similar results when subjects judged whether ellipses were the same height and whether color patches were the same hue, and Crowder (1989) also reported a similar result when subjects judged the pitches of tones that randomly varied in timbre. These results from the two-by-two orthogonal classification task are consistent with the suggestions by Fletcher and Rabbitt (1978) and Krueger and Shapiro (1981) that subjects bypass classification of the stimulus in favor of detection of trial-to-trial change. However, results from Experiments 2 and 3 (which use a more complex stimulus set) are not consistent with these suggestions. The results also are inconsistent with Felfoldy (1974) and Fletcher (1981), who reported that RTs are slowed by increased trial-to-trial differences between stimuli. The inconsistencies might be due to the particular procedures that were used in the latter two cases: limited practice on any condition (Felfoldy, 1974) and use of nonintegral state-limited dimensions (Fletcher, 1981). Fletcher's stimuli were numerals partially masked by a pattern of dots. The judged dimension was numeral identity, whereas the irrelevant dimension of dot number (5 or 10) limited the amount of information about the numerals that could be seen. This situation is different from when all information is potentially available to the subject (Garner, 1970) and probably should not be considered with studies in which all dimensional information is equally available to the subjects. In the Felfoldy study, the same stimuli were used

for relatively few trials in each of many different tasks during a single session. We think this encouraged subjects to use an object identification strategy, rather than to process attributes differently in different tasks as instructed. We attempted to obviate this possibility in two ways. We provided considerable practice, and we conducted a 2×12 condition in which identification of each of the 24 stimuli was unlikely (Miller, 1956). We consistently observed sequence effects different from those reported by Felfoldy (1974).

3. While sequence effects are associated with range, not all effects of range are due to sequence, as far as we have been able to measure. In Experiment 1, as stimulus range increased, and thus the magnitude of trial-to-trial difference increased, RTs on relevant-dimension change trials increased. But, RTs on irrelevant-dimension repetition trials also increased. This means that change on the irrelevant dimension between trials does not completely explain the observed range effects; instead, an overall slowing is present across all sequences, regardless of whether the irrelevant dimension changed. This is either because of some set effect (cf. Garner, 1962, chap. 5) in addition to sequence effects, or it is because our measure of sequence effects was incomplete, which may be possible, since sequence effects can extend further back than one trial (e.g., Staddon, King, & Lockhead, 1977). However, intervening stimuli frequently reduce effects of previous stimuli (Pashler & Baylis, 1991; E. E. Smith et al., 1973).

4. There were within-task range effects. These were reasonably expected given between-task range effects, but they had not previously been reported for orthogonal classification tasks. In Experiments 2 and 3, the RT difference between response-repetition trials and response-change trials was larger for larger between-trial changes on the irrelevant dimension. This within-conditions range effect was present in every condition and was relatively large. In Experiment 2A, which used the smallest range, there was, numerically, a more than 90-msec effect of range, even though no range effect was significant when only irrelevant-change trials were considered. This was because of the interaction between the effect when there was and when there was not irrelevant change between trials. These results preclude any complete account of classification performance that involves only detection of the presence or absence of change. Whether or not the relevant dimension changes and the magnitude of irrelevant trial-to-trial change are jointly important in determining RTs.

5. For the stimulus sets examined here, set composition did not affect range or sequence effects. RTs were very similar in Experiments 3A and 3B. Despite outlier stimuli in Experiment 3A that were not present in Experiment 3B, the RT slopes associated with response repetitions and response changes were highly similar, as were the lines' points of intersection. Also, RTs to outlier stimuli were similar to those to centrally clustered stimuli

when sequence was controlled. The only noted difference between these two data sets is that responses were slightly faster overall in Experiment 3B, and there is not enough information here to know whether this was due to the different stimulus sets or to the use of different subjects.

6. There was an important interaction. There were large irrelevant-dimension range effects (within a single condition) when the relevant dimension repeated. The slopes of RT as a function of irrelevant-dimension change (for response repetitions) were positive, reliable, and large (12.7 msec/step in Experiment 3A, 13.1 msec/step in Experiment 3B). However, there were very small (or nonexistent) irrelevant-dimension range effects when the relevant dimension changed. Then, the slopes of the lines representing RT as a function of the amount of irrelevant range were essentially zero. In Experiment 3A, this slope was -1.1 msec/step (see Figure 8); in Experiment 3B, it was -3.0 msec/step (see Figure 9). We conclude that the irrelevant dimension has no measured effect on trials in which the relevant-dimension level changes but has a large effect on trials in which the relevant-dimension level repeats.

These data are inconsistent with a bypass rule explanation (Fletcher & Rabbitt, 1978). Following the suggestions of Krueger and Shapiro (1981) for a *same-different* task, RTs to response changes should be speeded by increasing change between stimuli, since the likelihood of detection of that change also increases. But, here, the magnitude of the trial-to-trial change did not affect RT. So, the subjects were not simply using the detection of change from the previous trial to bypass a dimensional judgment process.

The Interpretation of "Filtering"

A frequent finding, replicated here, is that classification times, averaged over trials, are slower in orthogonal classification tasks than in univariate classification tasks. This difference has been interpreted as being due to analysis difficulties that subjects have when they are presented irrelevant information along with relevant information in a sorting task. The idea is that subjects must "filter out" the irrelevant information in order to classify the relevant information. This inference is sufficiently accepted such that this task—which was originally called *orthogonal sorting* to describe the fact that manipulations on the two stimulus dimensions are independent of one another (Garner, 1974, p. 124)—has become known, instead, as a *filtering* task. The name was changed to describe a theoretical interpretation of how irrelevant information makes it difficult for people to process relevant information: People must filter out the irrelevant information. At least for the integral stimuli that have been studied here, we question this interpretation.

The measure that is ordinarily used in reporting orthogonal classification data is the RT for a judgment of the relevant stimulus level, averaged across trials. This

summary measure masks any sequential information and produces the outcome that is misinterpreted as indicating filtering. Performance is not slowed because the subjects must filter out the varying irrelevant dimension in order to directly judge the relevant dimension. If it were, then there is no reason to predict that sequential behavior depends on whether or not the relevant dimension changed between trials, as it does. Rather, performance depends on sequence, at least in large part. Because there are different sequential combinations in control (univariate) tasks than in orthogonal (bivariate) tasks, performance is different in the two conditions.

Consider when the response changes between trials. Then, there is little to no effect on performance if the irrelevant level also changes, regardless of the magnitude of that change. There is no interference (see Figures 6–9). At least here, filtering out irrelevant information was not a problem to be understood since there was no interference to be accounted.

However, when the response repeats between trials, then there is an effect of irrelevant change. Furthermore, the magnitude of this slowing effect increases with increases in the magnitude of irrelevant change. Filtering might be involved here, but such an interpretation would mean that whether or not filtering is necessary depends on sequence. This is problematic since there is no a priori reason to suppose that whether or not the current stimulus must be decomposed depends on the previous trial.

The discrepancy between these two findings—that irrelevant variation both matters and does not matter within the same data set—indicates that something other than a general filtering difficulty is involved. We suggest that the results usually attributed to filtering are, instead, due to three different effects (noted next) that now need to be understood in order to create a better theoretical account of classification processing.

First, an overall set effect must be investigated and better understood. This effect appears in the data as an overall slowing of performance when the experiment is changed beyond the univariate procedure. This factor might be due to the requirement for many-to-one mappings (orthogonal sorting) rather than one-to-one mappings (univariate control tasks), which may make the overall task more difficult (Experiment 1), perhaps because subjects then must consider additional stimulus possibilities. Invoking some generalized filtering as the account is not adequate because it cannot account for the effects associated with range. More likely, we suggest, is that increasing either range or number of stimuli increases what must be attended to, and a greater attention demand is somehow associated with decreased quality of performance with any member of the total set.

Second, the sequential component associated with similar successive stimuli needs to be understood. Responses are slowed on response-change trials when there is a very small change in the irrelevant dimension (cf. Figures 8

and 9). This might be due to a difficulty in discrimination when the repeated levels are highly similar. This could be tested by using only large irrelevant changes in a condition.

Third, and most important because of its large effect, we need to explain the large negative effect on performance when the relevant-dimension level repeats but the irrelevant level changes. This may be due to response-selection processes as suggested briefly ahead.

A Process Account of Garner Interference

“Garner interference” describes situations in which the classification of values of some relevant dimension takes longer when an irrelevant dimension varies between trials than when there is no such variation. Essentially all attempts to explain this, with Felfoldy (1974) excepted, have attributed the slowed performance to difficulties the subjects have in analyzing the bivariate stimulus so as to evaluate the value of the relevant dimension. This interpretation is sufficiently ingrained that the orthogonal (two dimensions vary independently of one another) sorting task is commonly known as the *filtering sorting* task.

We suggest a slightly different interpretation that is consistent with the fact that there is no filtering difficulty when there is no trial-to-trial variation: Subjects do not initially analyze each stimulus into its attributes and process each independently. Rather, subjects compare each stimulus with what went before it. When the two are the same, as when the stimulus repeats, then subjects do not have to identify the magnitude of the relevant attribute—they simply have to say the same thing they said before (i.e., repeat the previous response). There are twice as many such repetitions in the univariate tasks (half of the trials) used as control conditions as in the orthogonal tasks (one fourth of the trials), and so average performance is relatively fast in the control conditions.

Similarly, it might seem, if the successive stimuli are very different, there is no need for analysis—subjects simply change the response. That works fine in the control conditions. However, in the orthogonal tasks, it would produce 25% errors, which includes all trials in which the irrelevant level changed but the relevant one did not. This means that the subjects must then be cautious. We suggest that a changed response is prepared when a stimulus change is detected and that, during the period between preparation and execution, the stimulus is analyzed for its value on the relevant dimension. This analysis requires more time when the successive events are more different (Lockhead, 1992a), so more time is needed for withdrawal of the response change.

Sequence Analyses

Sequence effects are not generally measured in classification tasks. Indeed, sequence effects have been considered annoyances to be averaged away because they interfere with understanding the more important contributions

of individual stimuli to judgment (Stevens, 1975, p. 66). We suggest, instead, that sequence effects provide information about how stimulus items are processed under different contexts (e.g., when different preceding stimuli are held in memory) and that classification processes cannot be understood without this information (Lockhead, 1992b). We propose this is the case for both univariate and bivariate classification data, and we suggest there are instances when sequential information can inform us about the processes involved in classification.

Consider orthogonal sorting tasks in this regard. These tasks have attracted considerable attention for more than 25 years, but progress in accounting for the noted slowing of performance, relative to that in univariate tasks, has been disappointingly slow. Possibly, one reason is that the theoretical efforts have been based on averaged data, which mask information revealed by sequential measures. This possibility needs to be entertained because sequential processes clearly are responsible for many of the differences seen in averaged classification data. It may be that sequential and, perhaps, other microanalytic measures will help us understand what is involved in classification.

There is no reason not to report sequential measures. No information is lost in doing so. Also, sequential analyses can be simply combined to provide the average measures that are commonly reported. The reverse is not the case: Sequential measures cannot be recovered from averaged measures. A regular practice of providing sequential measures whenever processes are masked by averaging, as regularly occurs in orthogonal classification tasks, should be useful.

REFERENCES

- BAIRD, J. C., BERGLUND, B., BERGLUND, U., & LINDBERG, S. (1991). Stimulus sequence and the exponent of the power function for loudness. *Perceptual & Motor Skills*, *73*, 3-17.
- BEN-ARTZI, E., & MARKS, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics*, *57*, 1151-1162.
- BERTELSON, P. (1961). Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology*, *13*, 90-102.
- BERTELSON, P. (1963). S-R relationships and reaction time to new versus repeated signals in a serial task. *Journal of Experimental Psychology*, *65*, 478-484.
- BERTELSON, P. (1965). Serial choice reaction-time as a function of response versus signal-and-response repetition. *Nature*, *206*, 217-218.
- BROADBENT, D. E. (1971). *Decision and stress*. New York: Academic Press.
- CROWDER, R. G. (1989). Imagery for musical timbre. *Journal of Experimental Psychology: Human Perception & Performance*, *15*, 472-478.
- DIXON, P., & JUST, M. A. (1978). Normalization of irrelevant dimensions in stimulus comparisons. *Journal of Experimental Psychology: Human Perception & Performance*, *4*, 36-46.
- DURLACH, N., & BRAIDA, L. (1969). Intensity perception I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, *81*, 421-427.
- EICHELMAN, W. H. (1970). Stimulus and response repetition effects for naming letters at two response-stimulus intervals. *Perception & Psychophysics*, *7*, 94-96.

- ELLS, J. G., & GOTTS, G. H. (1977). Serial reaction time as a function of the nature of repeated events. *Journal of Experimental Psychology: Human Perception & Performance*, **3**, 234-242.
- FARELL, B. (1985). "Same"-"different" judgments: A review of current controversies in perceptual comparisons. *Psychological Bulletin*, **98**, 419-456.
- FELFOLDY, G. L. (1974). Repetition effects in choice reaction time to multidimensional stimuli. *Perception & Psychophysics*, **15**, 453-459.
- FLETCHER, B. C. (1981). Wholistic and analytic stimulus processing: The development of selective perceptual strategies. *Quarterly Journal of Experimental Psychology*, **33A**, 167-176.
- FLETCHER, B. C., & RABBITT, P. M. (1978). The changing pattern of perceptual analytic strategies and response selection with practice in a two-choice reaction time task. *Quarterly Journal of Experimental Psychology*, **30**, 417-427.
- GARNER, W. R. (1962). *Uncertainty and structure as psychological concepts*. New York: Wiley.
- GARNER, W. R. (1970). The stimulus in information processing. *American Psychologist*, **25**, 350-358.
- GARNER, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.
- GARNER, W. R. (1978). Aspects of a stimulus: Features, dimensions, and configurations. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 99-133). Hillsdale, NJ: Erlbaum.
- GRAU, J. W., & KEMLER-NELSON, D. G. (1988). The distinction between integral and separable dimensions: Evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology: General*, **117**, 347-370.
- GRAVETTER, F., & LOCKHEAD, G. R. (1973). Criterial range as a frame of reference for stimulus judgment. *Psychological Review*, **80**, 203-216.
- HINRICH, J. V., & KRAINZ, P. L. (1970). Expectancy in choice reaction time: Anticipation of stimulus or response? *Journal of Experimental Psychology*, **85**, 330-334.
- HINSON, J. M., & LOCKHEAD, G. R. (1986). Range effects in successive discriminations. *Journal of Experimental Psychology: Animal Behavior Processes*, **12**, 270-276.
- HOLLAND, M. K., & LOCKHEAD, G. R. (1968). Sequential effects in absolute judgments of loudness. *Perception & Psychophysics*, **3**, 409-414.
- HYMAN, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, **45**, 188-196.
- KING, M. C., GRUENEWALD, P., & LOCKHEAD, G. R. (1978). Classifying related stimuli. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 417-427.
- KORNBLUM, S. (1973). Sequential effects in choice reaction time: A tutorial review. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 259-288). New York: Academic Press.
- KRUEGER, L. E., & SHAPIRO, R. G. (1981). Intertrial effects of same-different judgments. *Quarterly Journal of Experimental Psychology*, **33A**, 241-265.
- LI, X., & SMITH, A. F. (1992). Unitary classification in a comparison task. *Perception & Psychophysics*, **51**, 257-266.
- LOCKHEAD, G. R. (1966). Effects of dimensional redundancy on visual discrimination. *Journal of Experimental Psychology*, **72**, 95-104.
- LOCKHEAD, G. R. (1972). Processing dimensional stimuli: A note. *Psychological Review*, **79**, 410-419.
- LOCKHEAD, G. R. (1992a). On identifying things: A case for context. In B. Burns (Ed.), *Percepts, concepts, and categories: The representation and processing of information* (pp. 109-143). New York: Elsevier, North-Holland.
- LOCKHEAD, G. R. (1992b). Psychophysical scaling: Judgments of attributes or objects? *Behavioral & Brain Sciences*, **15**, 543-601.
- LOCKHEAD, G. R., GRUENEWALD, P., & KING, M. C. (1978). Holistic vs. attribute repetition effects in classifying stimuli. *Memory & Cognition*, **6**, 438-445.
- LOCKHEAD, G. R., & KING, M. C. (1983). A memory model of sequential effects in scaling tasks. *Journal of Experimental Psychology: Human Perception & Performance*, **9**, 461-473.
- LUCE, R. D., & GREEN, D. M. (1978). Two tests of a neural attention hypothesis for auditory psychophysics. *Perception & Psychophysics*, **23**, 363-371.
- MARKS, L. E. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 384-394.
- MELARA, R. D., & MARKS, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics*, **48**, 169-178.
- MELARA, R. D., & MOUNTS, J. R. W. (1994). Contextual influences on interactive processing: Effects of discriminability, quantity, and uncertainty. *Perception & Psychophysics*, **56**, 73-90.
- MELARA, R. D., & O'BRIEN, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, **116**, 323-336.
- MILLER, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, **63**, 81-97.
- PARDUCCI, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, **72**, 407-418.
- PARDUCCI, A., & PERRETT, L. F. (1971). Category rating scales: Effects of relative spacing and frequency of stimulus values. *Journal of Experimental Psychology*, **89**, 427-452.
- PASHLER, H., & BAYLIS, G. (1991). Procedural learning: 2. Intertrial repetition effects in speeded choice tasks. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 33-48.
- POLLACK, I. (1953). The information of elementary auditory displays. *II. Journal of the Acoustical Society of America*, **24**, 745-749.
- POMERANTZ, J. R. (1991). The structure of visual configurations: Stimulus versus subject contributions. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure* (pp. 195-210). Washington, DC: American Psychological Association.
- POMERANTZ, J. R., & GARNER, W. R. (1973). Stimulus configuration in selective attention tasks. *Perception & Psychophysics*, **14**, 565-569.
- POMERANTZ, J. R., PRISTACH, E. A., & CARSON, C. E. (1989). Attention and object perception. In B. Shepp & S. Ballesteros (Eds.), *Object perception: Structure and process* (pp. 53-89). Hillsdale, NJ: Erlbaum.
- RABBITT, P. M. (1968). Repetition effects and signal classification strategies in serial choice-response tasks. *Quarterly Journal of Experimental Psychology*, **20**, 232-240.
- SHEPARD, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, **1**, 54-87.
- SMITH, E. E., CHASE, W. G., & SMITH, P. G. (1973). Stimulus and response repetition effects in retrieval from short-term memory: Trace decay and memory search. *Journal of Experimental Psychology*, **98**, 413-422.
- SMITH, E. E., & MEDIN, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.
- SMITH, M. C. (1968). Repetition effect and short-term memory. *Journal of Experimental Psychology*, **77**, 435-439.
- STADDON, J. E. R., KING, M. C., & LOCKHEAD, G. R. (1977). On sequential effects in absolute judgment experiments. *Journal of Experimental Psychology: Human Perception & Performance*, **6**, 290-301.
- STEVENS, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects* (G. Stevens, Ed.). New York: Wiley.
- ST. JAMES, J. D., & ERIKSEN, C. W. (1991). Response competition produces a "fast same effect" in same-different judgments. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure* (pp. 157-168). Washington, DC: American Psychological Association.
- TEGHTSOONIAN, R. (1973). Range effects in psychophysical scaling and a revision of Stevens' law. *American Journal of Psychology*, **86**, 3-27.
- VAN LEEUWEN, C., & BAKKER, L. (1995). Stroop can occur without Garner interference: Strategic and mandatory influences in multidimensional stimuli. *Perception & Psychophysics*, **57**, 379-392.
- WILLIAMS, J. A. (1966). Sequential effects in disjunctive reaction time: Implications for decision models. *Journal of Experimental Psychology*, **71**, 665-666.

(Manuscript received August 11, 1997;
accepted for publication August 29, 1998.)