Memory & Cognition
1996, 24 (5), 557-572

The perception of number from the separability of the stimulus: The Stroop effect revisited

DANIEL ALGOM, AMNON DEKEL, and AINAT PANSKY
Bar-Ilan University, Ramat Gan, Israel

The literature on numerical perception is reviewed from the standpoint of research on selective attention, and predictions are made concerning the dimensional interaction between physical and numerical size of numerals. We manipulated stimulus differences to make the classification of numerical value slightly better (Experiment 1), substantially better (Experiment 2), or worse (Experiments 3–4) than classification of physical size. Garner, Stroop, and redundancy effects were used to gauge the degree of interactive processing. For nearly matched discriminability, both number and size appeared separable when the dimensions were varied orthogonally, but showed Stroop interference and redundancy gain when the dimensions were varied in a correlative fashion. When mismatched, asymmetric Garner and Stroop effects emerged in orthogonal contexts along with Stroop and redundancy gains in correlative contexts. These findings define a unique relation: Numerical value and physical size were optionally separable dimensions. We conclude that a magnitude representation is not mandatory for the perception of numerals. Our conclusions offer a new perspective for understanding both numerical perception and the Stroop phenomenon itself.

Seldom do we appreciate the paramount importance of numbers in our cognitive milieu. A number stands for our height, another for our IQ, a third indicates the balance in our checking account, a fourth our telephone—and the list is virtually inexhaustible. The ease with which we use numbers masks the fact that complex cognitive processes are required for the mere recognition of numerical stimuli or the execution of the simplest numerical comparisons and calculations. Tapping those processes poses a challenge for students of numerical cognition. In the present study we propose to use selective attention as a tool to uncover the nature of numerical perception. Simultaneously, we will use the results to reflect on models of selective attention itself. The classic measure of selective attention (indeed, its failure), of course, is that based on the seminal work of Stroop (1935), and Stroop-like effects have been reported for numerical dimensions. However, recent research has shown Stroop interference to be malleable, its presence depending on several heretofore neglected experimental factors. Therefore, despite a substantial amount of empirical study, we still lack a wide theoretical understanding of the Stroop effect, and we have notoriously little insight into processes of numerical cognition.

This research was supported by Israel Science Foundation Grant 58-93 to D.A. We thank Helena Kadlec, Robert Melara, Michael Masson, and Lawrence Marks for many insightful comments that improved the manuscript. Bob Melara inspired D.A.'s interest in selective attention when they both enjoyed the hospitality and scholarship of Larry Marks in the Pierce Laboratory at Yale University. We appreciate their continuing interest and support. Thanks are also due to Yoav Arieh, Joseph Glickson, and Lilach Shalev for detailed comments on earlier drafts. Correspondence should be addressed to D. Algom, Department of Psychology, Bar-Ilan University, Ramat Gan 52900, Israel.

In the present study, we use another method to assess selective attention to numerical dimensions, Garner's (1974) speeded classification paradigm. Garner's approach offers a broader framework to study numerical processing, and it permits one to derive the Stroop effect (if there is one) along with its own unbiased measures of selective attention. To anticipate our conclusions, selective attention research can reveal much about how we process numbers, and, consequently, about how best to model those processes. Our results yield new insights on both, and they challenge accepted models of numerical cognition. They also provide a potential perspective for a novel understanding of the Stroop phenomenon.

Numbers as Analog Magnitudes

The first study to measure the time needed to identify the larger of two single digits (Moyer & Landauer, 1967) yielded surprising results: The larger the numerical difference between the two digits, the shorter the time required to decide which is the larger. Regardless of years of experience or mathematical education, it takes longer to decide that 8 is larger than 6 than to decide that 8 is larger than 2. This *distance effect* has since been reproduced in numerous studies (see, e.g., Banks, Fujii, & Kayra-Stuart, 1976; Besner & Coltheart, 1979; Besner, Grimsell, & Davis, 1979; Buckley & Gillman, 1974; Dehaene, 1989; Dehaene, Dupoux, & Mehler, 1990; Duncan & McFarland, 1980; Foltz, Poltrock, & Potts, 1984; Garner, Podgorny, & Frasca, 1982; Henik & Tzelgov, 1982; Hinrichs, Yurko, & Hu, 1981; Jaffe-Katz, Budescu, & Wallsten, 1989; Katz, 1980; Link, 1990; Parkman, 1971; Poltrock & Schwartz, 1984; Restle, 1970; Sekuler, Rubin, & Armstrong, 1971; Takahashi & Green, 1983; Tzelgov, Meyer, & Henik, 1992).

Moyer and Landauer (1967) suggested that "the displayed numerals are converted to analogue magnitudes, and the comparison is then made between these magnitudes in much the same way that comparisons are made between physical stimuli such as loudness or length of line" (p. 1520). The perceptual comparator notion has informed models of numerical cognition that differ in nearly every other respect (see, e.g., Buckley & Gillman, 1974; Dehaene et al., 1990; Dehaene & Mehler, 1992; Gallistel & Gelman, 1992; Link, 1990; Restle, 1970; Sekuler et al., 1971; Shepard, Kilpatric, & Cunningham, 1975). Perhaps the strongest version of the analog models is that recently espoused by Gallistel and Gelman (1992), who have argued for the existence of a "preverbal system [that] uses magnitudes to represent numerosities" (p. 71). In the Gallistel and Gelman model, numbers are intimately linked to internal quantities or magnitudes in animals, children, and adults alike. The analog model is also commensurate with the results of several studies (e.g., Banks et al., 1976; Buckley & Gillman, 1974; Dehaene, 1989; Moyer & Landauer, 1967; Shepard et al., 1975) that show numerical judgments to obey Fechner and Weber's laws (but see Link, 1990; Sekuler & Mierkiewicz, 1977). That Fechner and Weber's laws operate in numerical comparisons has often been interpreted to underscore the perceptual nature of such comparisons.

Automatic Activation of Magnitude

Applications of the perceptual analog model often entail a further assumption that bears on the questions posed at the outset of this study. According to this assumption, not only do numbers map on to the referent magnitudes or quantities, but they do so *automatically*. Thus, Dehaene (1992) has claimed that "arabic numerals may rapidly and automatically evoke the internal quantity code" (p. 21), and Sudevan and Taylor (1987) have asserted that "the magnitude of the digit may evoke a response through a fast, automatic process over which the subjects have little control" (p. 94). Henik and Tzelgov (1982) have concluded that, in comparative judgments, "numerical distances are automatically computed" (p. 394), and Tzelgov et al. (1992) have repeatedly referred to the "autonomous processing of numerical information" (p. 167), because "when an object is perceived all its features are activated" (p. 178). Obligatory activation of certain numerical and arithmetical information is also implied by the work of LeFevre, Bisanz, and Mrkonjic (1988) and Zbrodoff and Logan (1986). The contention that "the presentation of an arabic numeral elicits an automatic activation of the appropriate . . . magnitude code," a code that "cannot be repressed, even though magnitude information is irrelevant to the task" (Dehaene, 1992, p. 21; see also Dehaene & Mehler, 1992), is implicit in various models of numerical processing (see, e.g., Gallistel & Gelman, 1992; Link, 1990; see also Shepard et al., 1975).

Therefore, the great bulk of the literature on numerical cognition entails the following two presumptions. First, numbers are mapped on to the same magnitude mode that underlies the processing of physical stimuli. Moreover, that magnitude code is contained ineluctably in representations of number. A corollary of those propositions is that numerical magnitude and physical size of the presented numerals comprise Stroop-like dimensions that interact in processing. Thus, variation in *physical* magnitude (or quantity) of the number stimuli should affect numerical perception and, conversely, variation in numerical magnitude should affect the processing of size or quantity. (The two physical codes often coincide, as in the prototypic case of the histogram, where magnitudes and numerosities are isomorphic; see Gallistel & Gelman, 1992). The implication is implicit in many studies of numerical perception, but it is explicit, for example, in the analog model suggested by Foltz et al. (1984) who, assuming that numerical comparisons are "similar to physical size comparisons" (p. 452), have inferred that the two processes "may interfere with one another," similar to the interference "observed in the Stroop task when a color name and a physical color conflict" (p. 453).

Stroop Interference With Numerical Dimensions

The general approach to testing Stroop-like interference in numerical perception can be illustrated by an experiment reported by Besner and Coltheart (1979). Subjects were shown pairs of arabic numerals and were timed on decisions about which of the two numerals was larger in value. On some trials, the numerically larger numeral was physically larger (congruent stimuli); on some trials the two number stimuli were the same physical size (same size, or neutral stimuli); on the remaining trials, the numerically larger numeral was physically smaller (incongruent stimuli). Besner and Coltheart found facilitation for congruent stimuli (i.e., yielding faster RTs than those obtained for neutral stimuli), and interference for incongruent stimuli (yielding longer RTs than those for neutral stimuli). Essentially the same results were obtained in other studies using similar designs: Stimuli mismatched for numerical and physical size resulted in slower RTs than either control stimuli (of the same physical size) or stimuli matched for numerical and physical size (e.g., Foltz et al., 1984; Hatta, 1983; Henik & Tzelgov, 1982; Takahashi & Green, 1983; Tzelgov et al., 1992; Vaid, 1985; Vaid & Corina, 1989). Parallel findings have been reported for numerosity (rather than physical magnitude) as the irrelevant dimension: Subjects suffered interference and reaped gain, respectively, for stimuli that mismatched and matched in numerical value and the number of numerical symbols in the display (see, e.g., Flowers, Warner, & Polansky, 1979; Morton, 1969; Shor, 1971).

Failures of selective attention have been found for judgments of physical size and numerosity as well (see, e.g., Flowers et al., 1979; Fox, Shor, & Steinman, 1971; Henik & Tzelgov, 1982; Hock & Petrasek, 1973; Morton, 1969; Reisberg, Baron, & Kemler, 1980; Shor, 1971; Tzelgov et al., 1992; Windes, 1968). Thus, not only did incongruent physical size interfere with judgments of numerical

size, but also, incongruent numerical size interfered with judgments of physical size or numerosity. The results of these studies suggest an inability to ignore numerical information even when it is irrelevant. Therefore, in general, the overall outcome of Stroop investigations of numerical dimensions seems to be consistent with the notion of obligatory activation of numerical information.

Unfortunately, Besner and Coltheart's (1979) study can also serve to illustrate the flaws and pitfalls associated with various Stroop investigations of numerical cognition. First, there was a glaring asymmetry in the number of stimuli used for the numerical and physical dimensions. Besner and Coltheart used the numbers 1 to 9 (inclusive) for the former, but only two values (2.3 × 1.7 cm ["large"] and 1.5 × 1.0 cm ["small"]) for the latter. This is typical: Virtually all pertinent research pitted a finely grained numerical dimension against an extremely coarse physical dimension. That asymmetry itself may have determined the magnitude and pattern of the observed interactions. Melara and Mounts (1994) have recently shown that the mere number of stimuli on an irrelevant dimension affects classification performance on the relevant dimension. Even more serious than the arbitrary selection of values for physical size, however, is that in none of the reviewed experiments was there an attempt to assess the *discriminability* of the numerical and physical dimensions.

Discriminability specifies the psychological difference separating two stimulus values along a dimension (Melara & Mounts, 1993). It is measured by the speed and accuracy needed to identify the two stimuli along a dimension as they alternate randomly from trial to trial. Note that discriminability is an intradimensional index, measured separately for each of the tested dimensions. Discriminability is matched if the values along the numerical dimension are as different psychologically as the values along the dimension of physical size. Of course, the different number of values used precluded the possibility of an overall match across the physical and numerical magnitudes. However, we do not have any indication of the discriminability of the two physical sizes that were used. This information is critical, because, as Melara and Mounts (1993) showed, Stroop interference is malleable, with the more discriminable dimension causing a failure of selective attention to the less discriminable dimension, but not vice versa. Thus, incongruent physical size will disrupt numerical performance if the values of size differ appreciably, but not if they are closer together. Conversely, incongruent numerical size will interfere with judgments of physical size only if the numbers are spread across a large range of values.

Consequently, given that discriminability has not been considered in earlier research, we question the generality of the conclusions drawn, and, in particular, the validity of the evidence adduced for the mandatory activation of numerical representation. We introduce a novel method to probe numerical interactions, one that does control for the relative discriminability of values along the tested dimensions. The speeded classification paradigm, developed by Garner (1974) and revitalized by Melara (Melara, 1989, 1992; Melara & Marks, 1990; Melara & Mounts, 1993, 1994; Melara & O'Brian, 1987) has been used extensively to study how irrelevant variation along one dimension affects performance on a second, relevant dimension. This paradigm is applied here to examine the existence and nature of the interaction(s) between physical and numerical magnitude of numbers.

Garner Interference With Numerical Dimensions

A single number stimulus is shown on each trial and the subject's task is to classify it on the relevant dimension as quickly as possible. For a numerical response, the subject is timed to classify two numbers as they alternate randomly from trial to trial. The numbers also differ along an irrelevant dimension, physical size, which the subject is instructed to ignore. Physical size constitutes the criterial dimension in a complementary condition, in which the subject is instructed to ignore the now-irrelevant variation in numerical value. Below, we describe the basic conditions constituting Garner's speeded-classification paradigm.

In the *filtering* conditions, subjects are asked to classify values on one dimension—say, whether numbers are 3s or 7s—while ignoring irrelevant variation on a second dimension—say, whether the physical size of the number is large or small. In the *baseline* conditions, subjects are again asked to classify values on the criterial dimension (e.g., whether numbers are 3s or 7s), but the values on the irrelevant dimension are held constant (e.g., all numbers are physically large). In the *correlated* conditions, the task again is merely to classify two stimuli on one dimension. However, the values of the stimuli on the two dimensions vary in a perfectly correlated manner throughout the sequence. Thus, the stimuli either match (e.g., all 3s are small and all 7s are large) or mismatch (all 3s are large and all 7s are small), respectively, for positively and negatively correlated sets of stimuli.

The ability to attend selectively is measured by comparing performance at baseline, where the irrelevant dimension is held constant, with performance in filtering, where the two dimensions are varied orthogonally. If performance in filtering equals that at baseline, selective attention is perfect. The parity implies that the subjects were able to focus on the criterial dimension (e.g., numerical value) without suffering distraction from irrelevant variation on another dimension (e.g., physical size). Conversely, if subjects' performance at classifying values on the relevant dimension is worse in filtering than at baseline, then selective attention has failed. The difference in RT between filtering and baseline is called *Garner interference* (Pomerantz, 1986), and it reflects an inability to focus exclusively on the relevant dimension. Pairs of dimensions that produce substantial Garner interference are called integral or *interacting* dimensions. Pairs of dimensions that do not lead to significant Garner interference are called *separable* dimensions.

Interactive and separable dimensions can be distinguished by further empirical criteria (Garner, 1974; Melara, 1992; Pomerantz, 1986). An important one relates to performance in the correlated tasks. For interactive dimensions, classification in the correlated tasks is better than that at baseline. For such continua, subjects reap gain from correlated variation of values on an irrelevant dimension (but suffer Garner interference from orthogonal variation). For separable dimensions, subjects neither reap redundancy gain nor suffer Garner interference, resulting in approximately equal performance at baseline, filtering, and correlated tasks.

Garner and Stroop Measures

Like Stroop's interference, Garner interference reflects a failure of selective attention. The two are separate indices of selective attention, yet both can be gauged under the same experimental context. The Stroop effect is gauged by comparing performance on congruent trials (in which numerical magnitude corresponds to physical magnitude) with performance on incongruent trials (in which numerical magnitude conflicts with physical magnitude). Three measures of Stroop, commensurate with this definition, ensue. First, in the filtering condition of the Garner paradigm, the irrelevant and relevant dimensions are varied orthogonally. Hence, on half the trials, numerical magnitude corresponds to physical magnitude, and on half, it conflicts. Therefore, Stroop effect equals the difference in classification RT between trials in which numerical value and physical size correspond and trials in which numerical value and physical size conflict. The same Stroop calculation applies to baseline. These measures have frequently served to assess Stroop within (e.g., Melara & Marks, 1990; Melara & Mounts, 1993, 1994; Pomerantz, 1983; Pomerantz, Pristach, & Carson, 1989) and independent of (MacLeod, 1991, and references therein) the Garner paradigm. A third measure of Stroop derives from the correlated tasks, defined as the difference in RT between positively and negatively correlated dimensions. This index is a between-conditions measure (as is Garner interference), whereas the former two are within-condition measures. We label the within-condition Stroop measures "Stroop congruity," and the between-condition Stroop measure "Stroop interference." We derived both Garner and the various Stroop effects on the numerical continua tested.

The Present Experiments

We recounted the substantial amount of evidence suggesting an analog model for numerical representations. The recurrent findings of a Stroop interference with numbers also are consistent with a mandatory magnitude representation. However, in earlier studies, researchers have neither tried nor appreciated the need to match discriminability on the tested dimensions. Information on discriminability is indispensable, however, because the very presence of a Stroop (and Garner) effect depends in part on an asymmetry in baseline discriminability. Matching discriminability may render the dimensions separable, and eliminate completely Stroop and Garner interference (Melara & Mounts, 1993; see also Garner & Felfoldy, 1970; Pomerantz, 1983). In fact, when Melara and Mounts (1993) matched discriminability on the original Stroop dimensions of word and color, they appeared to be separable. Traditional Stroop and Garner effects appeared only when the dimensions mismatched, with the more discriminable dimension disrupting classification on the less discriminable dimension. Indeed, Melara and Mounts concluded that the Stroop phenomenon is an optional effect, reflecting a failure of selective attention caused by unequal discriminability.

A series of pilot experiments (Algom, Dekel, & Pansky, 1993), in which we attempted to match the discriminabilities of numerical and physical magnitude of numerals, yielded just that result. Subjects could attend selectively to one numerical dimension in the face of irrelevant variation in the other numerical dimension. This outcome is incompatible with those of earlier studies, investigations that did not take into consideration the relative discriminabilities of the dimensions. Theoretically, if numerical separability proves general, it poses a serious challenge to analog accounts of numerical cognition.

Therefore, in this study, our goal was to explore the separability (or integrality) of numerical dimensions by evaluating selective attention to numerical and physical magnitude of numerals. We used Garner's speeded-classification paradigm and measured failure of selective attention through Garner and Stroop interference. Following Melara and Mounts (1993), we also manipulated discriminability. The results elucidate the multiple roles played by numbers in human information processing.

EXPERIMENT 1

The first experiment was aimed at replicating and extending the findings of Algom et al. (1993). Those results implied, surprisingly, that numerical dimensions might best be considered separable dimensions. In this experiment, we paired the numerical values 3 and 7 with two values of physical size. We tested these dimensions in the full speeded-classification paradigm—baseline, filtering, and correlated dimensions tasks. The present experiment was preceded by extensive preliminary testing to try to match the baseline discriminabilities of numerical value and physical size. In this way, we planned to provide for a fair test of selective attention to these two dimensions of number.

Method

Subjects. Twenty young men and women, mostly students recruited from the Bar-Ilan University community, were paid to participate.

Stimuli and Apparatus. Each number (3, 7) appeared at each physical size (small, large) to form four stimulus combinations. From these stimuli, we created 10 experimental tasks (5 involving

numerical magnitude classification, 5 involving physical magnitude classification) of 40 trials each. Thus, for numerical classification, the subjects performed in two baseline tasks (with physical size held constant at either large or small value), in two correlated dimensions tasks (with the numerical and physical values correlated, either positively or negatively), and in a filtering task (with physical size varying orthogonally). The subjects also performed in 5 complementary tasks, classifying physical size. A brief description of the task follows.

1. During baseline, subjects discriminated between two values on one dimension (e.g., 3 vs. 7), while the other dimension was held at a constant value (e.g., large).

2. In the filtering task, subjects were presented with all four stimuli, alternating randomly from trial to trial, and classified them as being either numerically large (7) or small (3), or, in a different block, as being either physically large or small.

3. In the positively correlated dimensions task, as in baseline, the stimulus set contained only two stimuli; by contrast, however, the set comprised the two congruent stimuli (i.e., small 3 and large 7). The subjects classified either numerical value or physical size.

4. In the negatively correlated dimensions task, subjects classified values from the two incongruent stimuli, again according to either numerical or physical magnitude.

From pilot testing it was determined that, to approximately equate speeds of numerical and physical discrimination, the numerals needed to measure 45 × 30 pixels (large) and 18 × 12 pixels (small) on the computer screen. The stimuli were generated in Pascal-small font by an IBM-compatible (PC-386) microcomputer and displayed on a super-VGA color monitor. The stimuli appeared white over a dark gray screen background near the center of the display; to avoid adaptation and other unwanted effects, we introduced trial-to-trial spatial uncertainty of 20 pixels around the center of the display. The subjects sat approximately 80 cm from the center of the screen. Equal numbers of stimuli were presented in the different tasks.

Procedure. The subjects were tested individually in a dimly lit room. The subjects performed the five numerical or five physical classifications together as a set, with half of the subjects first performing numerical tasks and half first performing physical tasks. Within each set, order of testing was varied in a semirandom fashion. Prior to performing a particular set, the subjects performed half the trials of the entire set of stimuli as practice.

Trials were presented randomly within each task. The subject initiated the first trial by pressing any key on the computer keyboard. The stimuli followed 0.5 sec after each response. Classifications were made by pressing either a right- or left-hand key on the keyboard. The stimuli were response terminated. RT was measured in milliseconds using a software timer. The subjects were instructed to attend to the relevant dimension and ignore irrelevant variation. They were encouraged to respond quickly, but accurately. Trials containing latencies that exceeded 800 msec were excluded from the analyses.

We adopted a Type I error probability of .05 throughout the following analyses; hence, any result deemed significant meets that criterion.

Results

Mean RTs and proportion of errors for numerical and physical judgments at baseline, filtering, and correlated tasks appear in Table 1. Average baseline performance was 337 msec for numerical value discrimination and 352 msec for physical size discrimination. Obviously, the two judgments were unequally discriminable at baseline, our extensive preliminary testing (on other subjects) notwithstanding. Numerical judgments were made 15 msec faster than were physical judgments; the difference was significant [$F(1,19) = 6.73$, $MS_e = 327.92$]. However, collapsing data across the various conditions showed that the two types of judgment were within 9 msec of each other, a nonsignificant difference [$F(1,19) = 3.01$, $MS_e = 3{,}249.0$, $p = .10$]. We shall offer an explanation for our failure to obtain equal discriminability at baseline. However, as we shall also see, this inequality does not jeopardize a coherent interpretation of the results. The correlation between speed and accuracy averaged .20 across subjects and conditions; thus, we have little evidence for a tradeoff between the two measures in the present experiments.

Remarkably, the magnitude of Garner interference was small (a mere 5 msec for classification of physical size), if it existed at all (−1 msec for numerical discrimination); statistical analyses showed that the effect was insignificant for either type of judgment ($F < 1$ in both cases). Therefore, what is most revealing about Experiment 1 is that subjects could attend well to either numerical value or physical size without suffering intrusions from the other dimension.

Positively correlated dimensions produced a reliable redundancy gain: 12 msec for numerical judgments [$F(1,19) = 4.69$, $MS_e = 327.9$] and 34 msec for physical judgments [$F(1,19) = 36.61$, $MS_e = 327.9$]. Negatively correlated dimensions, by contrast, led to neither gain nor loss relative to baseline for either type of classification [$F(1,19) = 0.06$ and 2.42, $MS_e = 327.9$, $p > .10$, respectively, for numerical and physical classifications]. This pattern of data, namely substantial redundancy gain for positively correlated dimensions, but neither gain nor

Table 1
Mean Reaction Times (RTs) (in Milliseconds) and Proportions of Error for Judgments of Numerical and Physical Size, in Baseline, Filtering, and Correlated Dimensions Tasks (Experiment 1)

	Number				Size					
	RT		Error		RT		Error		Overall Mean	
Task	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	RT	Error
Baseline	337	43	.06	.04	352	41	.08	.04	344	.07
Filtering	336	35	.07	.04	357	42	.09	.06	346	.08
Positive correlation	325	36	.06	.04	318	37	.04	.03	321	.05
Negative correlation	336	43	.07	.03	343	40	.06	.04	339	.07
Overall mean	333		.07		342		.07		337	.07

interference for negatively correlated dimensions, replicates that first reported by Melara and O'Brian (1987) for cross-modally (synesthetically) corresponding dimensions. Note, however, that the present data produced no Garner interference, whereas those by Melara and his colleagues did. Melara (1989; Melara & Marks, 1990; Melara & O'Brian, 1987) has suggested distinguishing between *interacting* dimensions producing exactly that pattern and *integral* dimensions yielding data that conform fully to the original Garnerian definition of dimensional integrality.

A glance at Table 1 also shows that error data were consistent with the RT data, but that the effects were generally weaker. Although the error rates precisely mirrored the corresponding RTs, the only significant effect obtained was that for redundancy gain at positively correlated dimensions for judgments of physical size [$t(19) = 3.01$, $SE_{dm} = 1.43$]. To be sure, reaction times are generally more stable in speeded classification (see Melara & Marks, 1990; Melara & Mounts, 1993), wherefore we emphasize RTs in our discussion.

Table 2 presents our measure of Stroop congruity for the dimensions tested. Stroop congruity was defined as the difference in RT to congruent (matching) and incongruent (mismatching) stimuli in a task. A positive value indicates shorter RTs to congruent stimuli, that is, the presence of Stroop facilitation or interference. For each subject, we derived these Stroop scores at baseline and in filtering, and the means appear in Table 2.

A striking feature of the data shown in Table 2 is the minuscule amounts of Stroop congruity observed. Although the congruity scores are larger for physical judgments than for numerical judgments, none of the scores differs significantly from zero. Incongruent stimuli were responded to almost as quickly as were congruent stimuli, the near equality resulting in tiny amounts of Stroop congruity in both baseline and filtering. For these tasks, then, the pattern observed with Stroop largely corroborates those observed with Garner. Both measures show that the irrelevant dimension of numerical stimuli did not interfere strongly with processing of the relevant dimension.

The redundancy gains observed for positively correlated dimensions imply, by contrast, an interaction between numerical value and physical size. Moreover, the difference in RT between positively correlated dimensions and negatively correlated dimensions defined substantial Stroop effects: 11 msec for numerical judgments and 25 msec for judgments of physical size. Both the overall Stroop effect (18 msec) and the effect for physical judgments were significant [$t(19) = 2.11$, $SE_{dm} = 8.519$, and 2.20, $SE_{dm} = 11.36$, respectively], although that for numerical judgment was not [$t(19) = 0.89$]. Therefore, for correlated tasks as well, Garner and Stroop converge. Both measures showed that, in a context in which the values on an irrelevant dimension covaried positively with values on the relevant dimension, subjects reaped gain in classification performance.

Table 2
Stroop Congruity: Mean Reaction Times (in Milliseconds) to Congruent and Incongruent Stimuli for Judgments of Numerical and Physical Size in Baseline and Filtering Tasks (Experiment 1)

Task	Congruent		Incongruent		Congruity Score
	M	*SD*	*M*	*SD*	
Numerical size					
Baseline	340	40	343	42	+3
Filtering	335	39	337	34	+2
Physical size					
Baseline	346	40	354	43	+8
Filtering	350	51	357	45	+7

Discussion

The results of Experiment 1 suggest conclusions similar to those reached by Algom et al. (1993) for numerical and physical magnitude, and by Melara (e.g., Melara & Mounts 1993; Melara & Marks, 1990) for Stroop (color/word) stimuli in general: Subjects are able to attend selectively to numerical magnitude or physical magnitude without suffering intrusions from random variation on the unattended dimension. The absolute amount of interference, whether of Garner or Stroop, was minimal or nonexistent. The results imply that, when baseline discriminabilities are within 15 msec of each other, subjects can process the dimensions of numerical and physical magnitude quite independently. We conclude that the present numerical attributes constitute largely separable dimensions.

In earlier studies, researchers testing for interference effects in numerical cognition have ignored the relative discriminability of stimuli along the tested dimensions; we measured and approximately equated it here. The results of studies reviewed earlier suggest bidirectional interaction between the dimensions of numerical value and physical size; we found none here. Our results imply that numbers do not automatically call up representations of size. Consequently, we conclude, theories that posit mandatory activation of magnitude in numerical processing are probably misguided.

The substantial Stroop interference and redundancy gain obtained for the correlated tasks threaten to compromise our conclusion. These results are perplexing because they imply that numerical value and physical size interact, an interaction that is conspicuously absent in other portions of the data. Thus, the absence of Garner interference and Stroop congruity at filtering and baseline suggests that numerical value and physical size are separable dimensions. On the other hand, the redundancy gain and Stroop interference measured in correlated contexts suggest that numerical value and physical size are interacting dimensions. We discuss two general explanations of this unique pattern of data.

The first explanation stresses the primary role of the filtering task in determining integrality or separability. The reason is that, even with separable dimensions, it is often optimal for subjects to attend to the nominally irrelevant dimension in the positively correlated task. Hence, the correlated tasks provide a weak test of separability. Gauged by performance at filtering, numerical

dimensions are separable, and the classic probability theorem for independent observations suggests one possible explanation for the *entire* pattern of the results: The decrease in RT for positively correlated dimensions is due to an increase in the probability of reporting one of a series of independent events as the number of these events increases. In a positively correlated context, detecting either attribute of the congruent stimuli leads to a correct response, thereby maximizing performance. Thus, according to the probability summation hypothesis, performance in a positively correlated context improves not because there is an interaction between numerical and physical magnitude, but because of multiple opportunities to detect by *independent* channels (see Algom & Babkoff, 1984). Possibly, our observers ignored the irrelevant dimension in orthogonal contexts (to avoid interference), but paid attention to it in positively correlated contexts (to reap gain). No redundancy gain obtained in negatively correlated contexts, either because of response competition from the semantic conflict between values on the dimensions, or simply because observers decided to ignore the irrelevant dimension in those contexts (as they did in filtering).

A second and more radical explanation of the data entails differential semantic processing of the same stimuli in orthogonal and correlated contexts. According to this interpretation, in filtering, the subjects did not process the number stimuli semantically—qua numbers—but, instead, as shapes. Conceivably, our subjects ignored the nominal attributes stated in the instructions ("numbers"), and preferred to attend instead to other attributes (shapes) with no costs incurred. Pomerantz (1991) has argued that perceivers can voluntarily decide what features or dimension to attend, and that the latter do not necessarily coincide with those conveyed by the experimental instructions. If indeed the subjects were classifying the shapes 3 and 7 when instructed to classify numbers, the absence of Garner interference and Stroop congruity should not come as a surprise, because shape and size are prime examples of separable dimensions (see, e.g., Garner, 1974; Garner & Felfoldy, 1970). By contrast, in positively correlated contexts, subjects did attend to numbers, as attested to by the sizable amounts of redundancy gain and Stroop interference obtained. If numbers were processed as shapes in these contexts, neither effect would be expected, because, as shapes, 3 and 7 do not have a correspondence quality to share with values of size (and one cannot properly speak of congruent and incongruent stimuli). Of course, over the trials, subjects could notice the correlation between a particular shape and a particular size and respond more rapidly. However, in that case, one would expect a modest amount of gain for both positively *and* negatively correlated dimensions, and *no* Stroop interference, none of which occurred.

Let us issue a caveat before proceeding. Neither we nor Pomerantz argue that the selection of stimulus attributes induced by context comes under conscious control. By asserting strong contextual influences on semantic processing, we do not appeal to conscious deliberation on the part of the observer.

We reiterate, therefore, the conclusion stated at the outset of this discussion: Numerical value and physical size are separable dimensions, albeit ones that, under privileged contexts, behave in an integral fashion. On the basis of the present pattern of data, we propose to distinguish a unique type of separable dimensions, *optional dimensions*, characterized by the lack of Garner and Stroop effects in orthogonal contexts, but by redundancy gain and Stroop interference in correlated contexts. We will elucidate the broader implications of these data for numerical cognition on the one hand and selective attention on the other.

EXPERIMENT 2

In Experiment 2, we tested classification of numerical value and physical size when the baseline discriminabilities were markedly mismatched. We planned to reproduce the Stroop effects reported in the literature by making the numerical dimension much more discriminable than the physical dimension. To effect the mismatch, we presented physical sizes closer together in value than those used in Experiment 1. The manipulation caused a slowdown in size classification, but left classification of numerical value largely unchanged. Following Melara and Mounts (1993), we predicted a pattern commensurate with the usual Stroop asymmetry, with numerical value affecting size judgments, but size having no appreciable effect on judgments of numerical value.

Method

Subjects. Ten subjects, again paid volunteers from the Bar-Ilan community, participated in this experiment. None had participated in Experiment 1.

Stimuli, Apparatus, and Procedure. The stimuli and apparatus were identical to those in Experiment 1. The values of physical size used were 22 × 15 and 36 × 24 pixels, respectively, for the small and large formats. In all other respects, Experiment 2 duplicated Experiment 1. Again, in different sets of blocks, the subjects were instructed to classify numbers according to either numerical or physical size as quickly as possible.

Results

RTs and errors rates for numerical value and physical size classifications appear in Table 3. Baseline performance in number identification (319 msec) was 75 msec faster than baseline performance in physical size identification (394 msec) [$F(1,9) = 54.66$, $MS_e = 509.054$]. Collapsing data across all tasks, the two types of classification differed by 69 msec, with numerical value being the easier dimension [$F(1,9) = 44.04$, $MS_e = 94{,}393.8$]. Therefore, we were successful in mismatching discriminabilities in favor of numerical magnitude.

Next, consider the effects on classification performance of orthogonal variation of values along the irrelevant dimension. For classification of physical size, the filtering task (415 msec) was performed 21 msec more slowly on

Table 3
Mean Reaction Times (RTs) (in Milliseconds) and Proportions of Error for Judgments of Numerical and Physical Size, in Baseline, Filtering, and Correlated Dimensions Tasks (Experiment 2)

	Number				Size					
	RT		Error		RT		Error		Overall Mean	
Task	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	RT	Error
Baseline	319	33	.08	.04	394	58	.10	.04	356	.09
Filtering	321	41	.06	.03	415	62	.10	.03	368	.08
Positive correlation	316	49	.07	.03	365	48	.07	.03	340	.07
Negative correlation	337	40	.11	.09	394	68	.07	.04	365	.09
Overall mean	323		.08		392		.08		357	.08

the average than was the baseline task (394 msec). For this dimension, Garner interference was significant [$F(1,9) = 4.58$, $MS_e = 509.054$]. For classification of numerical value, by contrast, filtering was slower than baseline by 2 msec, an insignificant Garner interference ($F < 1$). Thus, the pattern of results differed for judgments of numerical value and physical size.

The error data largely mirrored those for RT, yet none of the differences was significant at the .05 level. Again, our interpretations emphasize the RT data.

Table 4 provides a summary of the Stroop scores at baseline and in filtering for both types of judgments. For numerical value, attributes from congruent stimuli were classified about as fast as were attributes from incongruent stimuli, yielding insignificantly small amounts of Stroop congruity. For physical size, attributes from congruent stimuli were classified 15 msec (baseline) and 25 msec (filtering) faster on average than were attributes from incongruent stimuli. Both measures of Stroop congruity were significant [$t(9) = 2.36$, $SE_{dm} = 6.54$, and $t(9) = 3.25$, $SE_{dm} = 7.69$, respectively]. Stroop congruity thus mirrors Garner's interference for both types of classification: Either both are substantial (physical size), or neither is (numerical value).

Finally, consider the effects of correlated dimensions (Table 3). These results, again, are intriguing. For judgments of physical size, stimuli in the positively correlated dimensions task were classified 29 msec faster than those at baseline, a significant redundancy gain [$F(1,9) = 8.26$, $MS_e = 509.054$]. In addition, the congruent stimuli in the former condition were classified 29 msec faster than were the incongruent stimuli in the negatively correlated dimensions condition, yielding an equally large measure of Stroop interference [$t(9) = 16.84$, $SE_{dm} = 1.722$]. Thus, the results for physical size judgments conform fully with an interacting pattern: Garner interference in an orthogonal context, redundancy gain in a positively correlated context, and neither gain nor loss in a negatively correlated context. Stroop congruity and interference matched perfectly, with significant effects in both orthogonal (baseline, filtering) and correlated contexts.

For classification of numerical value, we found no redundancy gain, relative to baseline, when irrelevant size varied in a corresponding fashion. Thus, positively correlated dimensions (316 msec) were classified a mere 3 msec faster than were baseline (319 msec), an insignificant gain ($F < 1$). However, the subjects incurred loss in the negatively correlated dimensions task: Their performance was slowed down by 18 msec relative to baseline [$F(1,9) = 3.04$, $MS_e = 509.054$, $p = .09$]. Perhaps more startlingly, subjects classified the incongruent stimuli (negatively correlated dimensions) 21 msec more slowly than they did the congruent stimuli (positively correlated dimensions), yielding a significant Stroop effect [$t(9) = 2.31$, $SE_{dm} = 9.09$]. Therefore, for judgments of numbers, the dimensions of numerical and physical magnitude were optionally separable: They produced neither Garner interference nor Stroop congruity in orthogonal contexts, but performance was harmed when the attributes conflicted and Stroop interference occurred in correlated contexts.

Table 4
Stroop Congruity: Mean Reaction Times (in Milliseconds) to Congruent and Incongruent Stimuli for Judgments of Numerical and Physical Size in Baseline and Filtering Tasks (Experiment 2)

	Congruent		Incongruent		Congruity
Task	*M*	*SD*	*M*	*SD*	Score
Numerical size					
Baseline	318	32	320	37	+2
Filtering	322	36	324	47	+2
Physical size					
Baseline	386	57	401	60	+15
Filtering	403	60	428	66	+25

Discussion

The results of Experiment 2 differ fundamentally from those of Experiment 1. Our explanation implicates baseline discriminability as the source of this difference. In Experiment 2, the inequality favoring numerical magnitude was five times the value in Experiment 1. That mismatch in favor of number identification yielded asymmetrical effects of Garner and Stroop. The less discriminable dimension (physical size) was affected by the more discriminable dimension (numerical value), but numerical value was largely unaffected by physical size. Judgments of size yielded considerable Garner and Stroop effects at both baseline and in filtering, as well as redundancy gain and Stroop interference in the correlated tasks. In contradistinction, judgments of numerical value were free of Garner and Stroop effects in the orthogonal context and produced no redundancy gain in the positively cor-

related context (but they suffered loss in the negatively correlated dimensions task and produced Stroop interference for the correlated context).

The results of Experiment 2 show that, by manipulating discriminability, we have successfully fabricated a Stroop effect that has been reported in the literature: The numerical value of the displayed numerals affects judgments of their physical magnitude or numerosity (see, e.g., Flowers et al., 1979; Henik & Tzelgov, 1982; Morton, 1969; Tzelgov et al., 1992; Windes, 1968). Nevertheless, viewed together with Experiment 1, our results imply that numerical dimensions do not lead inevitably to a breakdown of selective attention and impaired performance. On the contrary, our results render earlier conclusions suspect in that the observed interactions could easily have arisen from processing relative variation along the tested dimensions.

The results from the correlated dimensions clearly suggest that subjects were sensitive to the correspondences between numerical and physical magnitude. Thus, subjects were able to integrate numerical value and physical size when it was in their best interest to do so. They exercised that option in positively correlated contexts, demonstrating considerable strategic influences in the processing of numerical dimensions. In the current experiment, Stroop interference arose not because of a redundancy gain from positively correlated dimensions (Experiment 1), but rather because of a loss from negatively correlated dimensions. In general, sensitivity to the correspondences engendered by semantic processing of stimuli in correlated contexts may lead to performance gains, losses, or both gains and losses. Subjects may increase their speed for positively correlated dimensions (Experiment 1), decrease their speed for negatively correlated dimensions (Experiment 2), do both (Experiment 3, below), or increase speed for negatively correlated dimensions (Experiment 4, below). Whether these strategies entail further differences in the pattern of dimensional processing must await systematic study.

EXPERIMENT 3

The results from Experiment 1 suggested dimensional separability between numerical and physical magnitude when the former was easier by no more than 15 msec at baseline; neither Garner interference nor Stroop congruity emerged under those conditions. The results from Experiment 2 showed asymmetrical Garner and Stroop effects when the dimensions were greatly mismatched in favor of numerical value. Under the latter conditions, we reproduced the traditional Stroop asymmetry, with numerical value intruding on judgments of size but judgments of number largely unaffected by variation in size. In Experiment 3, we hoped to fabricate the reverse of the usual Stroop effect by making physical size the more discriminable dimension. We were only partly successful. Our failure might be helpful, however, in aiding in the diagnosis of the actual dimensions to which our subjects chose to attend.

Method

Subjects. Fourteen young men and women from the Bar-Ilan community were paid to participate. None had participated in either of the earlier experiments.

Stimuli, Apparatus, and Procedure. The apparatus and procedure were identical to those in Experiment 1. The physical sizes of the numerals were 13 × 9 and 72 × 48 pixels. Pilot work indicated that, at those values, judgments of size were faster than were judgments of numerical value. They constituted the smallest and largest values of size possible without compromising the nature of our stimuli (i.e., without making the small numeral illegible and the large one disparate in form). In fact, any increase in physical range threatened to disrupt identification of the small numeral and change the form of the large numeral.

Results

Baseline number discrimination (325 msec) was 10 msec slower than baseline size discrimination (315 msec), an insignificant difference [$F(1,13) = 2.28, p = .14$]. The overall mean for numerical classifications (330 msec) was 13 msec slower than the mean for physical classifications (317 msec) and this difference was reliable [$F(1,13) = 11.08, MS_e = 5,143.58$]. Thus, as Table 5 shows, discrimination was tilted in favor of physical size, but the inequality was not great. Our manipulation succeeded in speeding up the physical responses, but the advantage it engendered for those responses fell short of significance at baseline.

Next, consider the effects of uncorrelated dimensions. For classification of numerical value, the filtering task was performed 21 msec more slowly on average than was the baseline task, a significant effect of Garner interference [$F(1,13) = 11.46, MS_e = 273.12$]. For classification of physical size, by contrast, the filtering task was performed only slightly (6 msec) more slowly on average

Table 5
Mean Reaction Times (RTs) (in Milliseconds) and Proportions of Error for Judgments of Numerical and Physical Size, in Baseline, Filtering, and Correlated Dimensions Tasks (Experiment 3)

	Number				Size					
	RT		Error		RT		Error		Overall Mean	
Task	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	RT	Error
Baseline	325	43	.09	.04	315	41	.08	.03	320	.08
Filtering	346	39	.09	.04	321	51	.08	.03	333	.08
Positive correlation	315	44	.06	.03	313	44	.07	.05	314	.06
Negative correlation	335	39	.07	.04	318	51	.08	.07	326	.07
Overall mean	330		.08		317		.08		323	.08

than was the baseline task, an insignificant Garner interference ($F < 1$). Thus, random variation of values on the irrelevant dimension disrupted numerical classification, but left physical classification largely unimpaired. For these data, too, errors largely (though not fully) reproduced the pattern for RTs, although, again, none of the differences was statistically significant.

Table 6 presents the values of Stroop congruity at baseline and in filtering. For both numerical value and physical size, there were substantial effects of Stroop congruity in the filtering tasks [$t(13) = 2.49$, $SE_{dm} = 5.613$, and $t(13) = 4.16$, $SE_{dm} = 6.255$, respectively], but not at baseline [for numerical value, $t(13) = 1.87, p > .10$]. Thus, in Experiment 3, subjects responded more rapidly to congruent than to incongruent stimuli, even in orthogonal contexts. Notable, too, is the first appearance of a Stroop effect for numerical judgments in these experiments. However, a Stroop effect appeared for size judgments as well, despite their superior discriminability. The former but not the latter effect is commensurate with the mismatch in discriminability favoring physical size.

Finally, consider the results from the correlated dimensions tasks. For numerical discrimination, there was a 10-msec redundancy gain for positively correlated dimensions, and a 10-msec loss for negatively correlated dimensions. Although noticeable, neither effect was quite significant [$F(1,13) = 2.64$, and 2.60, $p = 0.11$ and 0.12 , respectively, for positively and negatively correlated dimensions]. However, the difference in RT between the two contexts defines a significant Stroop effect [$t(13) = 7.48$, $SE_{dm} = 2.674$]. For classification of size, by contrast, we found neither gain (2 msec) at positively correlated dimensions nor loss (3 msec) at negatively correlated dimensions relative to baseline (for both tasks, $F < 1$). Consequently, neither was there a significant Stroop effect [$t(13) = 1.51, p > .05$]. Thus, in correlated contexts, the pattern of results differed for numerical and physical discriminations. For the former, there were modest amounts of gain and loss and a significant Stroop effect; for the latter, more discriminable dimension, we did not find gain, or loss, or a significant Stroop interference.

Discussion

Although physical size was only slightly more discriminable than numerical value, an asymmetric pattern obtained, with the former dimension interfering with the latter dimension more pervasively than vice versa. Despite the suggestive results, our failure to create a greater advantage for size at baseline poses a problem vis-à-vis the broader conceptual framework of this study. In Experiment 1, at baseline, numerical value was more discriminable than physical size by a significant 15 msec, and the dimensions were largely separable. In Experiment 3, at baseline, physical size was more discriminable than was numerical value by an insignificant 10 msec, and an asymmetrical interaction between dimensions was obtained. Why did a slight mismatch in favor of numerical value leave the dimensions separable, while an even slighter mismatch in favor of physical size engendered interference? For a clue, one should consult the overall means for RT, not just those at baseline. The grand means for numerical and physical classifications did not differ significantly in Experiment 1, but they did in Experiment 3. Moreover, discrimination of physical size was faster than discrimination of number on all tasks in Experiment 3, but the RTs crossed in Experiment 1. Therefore, collapsed over the entire set of judgments, the data in Experiment 3 did differ more on the two dimensions.

Nevertheless, why was numerical value that resilient in the face of our efforts to render it the less discriminable dimension? A compelling explanation, again, is based on the aforementioned idea of Pomerantz (1991), suggesting that subjects have considerable latitude in selecting the attended features of the presented stimuli. Our subjects might have judged shapes when instructed to judge numbers (in orthogonal contexts). If they did, this would explain the extremely short RTs obtained for the number dimension. The physical shapes 3 and 7 are probably much more discriminable than physical size, an advantage that proved difficult to overcome. Logically (see Garner, 1974, 1983; but see Melara & Marks, 1990), the physical dimension of shape precedes that of size on a hierarchy of psychological levels of processing.

Despite the generally congruous results of Experiment 3, our failure to mismatch discriminability in favor of the size dimension might compromise a straightforward explanation of the data of this study taken as a whole. Consequently, in Experiment 4, we made another attempt to create a sizable mismatch in discriminability in favor of physical size.

Table 6
Stroop Congruity: Mean Reaction Times (in Milliseconds) to Congruent and Incongruent Stimuli for Judgments of Numerical and Physical Size in Baseline and Filtering Tasks (Experiment 3)

Task	Congruent		Incongruent		Congruity Score
	M	*SD*	*M*	*SD*	
Numerical size					
Baseline	320	44	330	43	+10
Filtering	339	41	353	39	+14
Physical size					
Baseline	317	41	314	41	−3
Filtering	308	44	334	59	+26

EXPERIMENT 4

In Experiments 1–3, we manipulated discriminability by changing values along the size dimension, but leaving the values along the number dimension unchanged. However, using 3 and 7 as stimuli, mismatching discriminability in favor of physical size proved difficult, probably because the subjects were classifying shapes in spite of instructions to classify numbers (in baseline and filtering). Therefore, in Experiment 4, we changed instead numerical magnitude and used the numerals 6 and 8, because these stimuli are closer together in numerical value

as well as in shape. The manipulation caused a slowdown in numerical classification relative to classification of size. Commensurate with a reverse Stroop asymmetry, in this experiment, physical size affected numerical judgments, but numerical magnitude did not exert an appreciable effect on judgments of size.

Method

Subjects. Fourteen young men and women from the Bar-Ilan community were paid to participate. None had participated in any of the earlier experiments.

Stimuli, Apparatus, and Procedure. We used 6 as the numerically smaller stimulus, and 8 as the numerically larger stimulus. The physical sizes of the numerals were 16 × 11 (small) and 128 × 88 (large) pixels, generated in a sans serif font. In all other respects, the apparatus and procedure were identical to those in Experiment 1.

Results

Mean RTs and proportions of errors for numerical and physical judgments in baseline, filtering, and correlated tasks appear in Table 7. Obviously, the two judgments were unequally discriminable at baseline, with size identification (276 msec) 32 msec faster than number identification (308 msec), [$F(1,13) = 42.17$, $MS_e = 170.72$]. Collapsing data across all tasks yielded an equally significant difference between the two types of classification, with physical size being the easier dimension [$F(1,13) = 40.03$, $MS_e = 21{,}694.46$]. Therefore, in Experiment 4, we were successful in mismatching discriminabilities in favor of physical size. The error data closely followed those for RT, underscoring once again the fact that the present results do not derive from a speed-accuracy tradeoff. Although accuracy did not differ significantly at baseline [$t(13) = 1.1$, $p = .29$], the few effects obtained for error (cited below) support our general conclusions.

For classification of numerical magnitude, the filtering task (327 msec) was performed 19 msec more slowly on the average than was the baseline task (308 msec). For classification of physical size, by contrast, filtering was slower than baseline by a mere 2 msec. Statistical analyses show that Garner interference was sizable and significant for numerical judgments [$F(1,13) = 14.00$, $MS_e = 170.72$], but it was absent for judgments along the easier size dimension ($F < 1$). Thus, physical size intruded on judgments of numerical size, but numerical size did not interfere with judgments of physical size.

Table 8 shows the Stroop scores at baseline and in filtering for the two types of judgments. For the easier dimension of size, Stroop congruity scores were small and insignificant at both baseline and filtering. For numerical judgments, by contrast, Stroop scores were significant at both baseline [$t(13) = 4.47$, $p < .001$] and filtering [$t(13) = 2.75$, $SE_{dm} = 4.727$]. Thus, both Stroop and Garner effects were substantial for classification of numerical value, but neither was substantial for classification of size. Notable are the negative values of Stroop congruity obtained for numerical size: Attributes from incongruent stimuli were classified 14 msec (baseline) and 13 msec (filtering) more rapidly on average than were attributes from congruent stimuli. We provide an explanation for this surprising finding.

To evaluate results from the correlated tasks, examine again Table 7. For classification of physical size, we found neither redundancy gain for positively correlated dimensions (279 msec) nor loss for negatively correlated dimensions (281 msec) relative to classification performance at baseline (276 msec). Obviously, too, the subjects did not classify incongruent stimuli (negatively correlated dimensions) more slowly than they did congruent stimuli (positively correlated dimensions), defining an insignificant (2-msec) Stroop interference ($F < 1$ for the three effects). For classification of numerical magnitude, the results differ. Although positively correlated dimensions did not produce a significant redundancy gain (6 msec, $F = 1$), negatively correlated dimensions did [21 msec, $t(13) = 2.17$, $SE_{dm} = 9.67$]. Consequently, a significant Stroop effect also resulted from a 17-msec difference in classification of congruent and incongruent stimuli, with the latter classified faster [$t(13) = 3.29$, $SE_{dm} = 5.167$]. For errors, we found significant redundancy gains for both positively and negatively correlated dimensions in numerical judgments [$t(13) = 3.87$, $SE_{dm} = 0.75$, and $t(13) = 3.99$, $SE_{dm} = 0.76$, respectively].

The one unexpected finding of Experiment 4 was the faster numerical classification of incongruent stimuli. However, the conundrum vanishes upon a detailed examination of the data: For one stimulus, 6, our definition of congruity did not conform to that espoused by the subjects. By experimental definition, a large 6 constitutes

Table 7
Mean Reaction Times (RTs) (in Milliseconds) and Proportions of Error for Judgments of Numerical and Physical Size, in Baseline, Filtering, and Correlated Dimensions Tasks (Experiment 4)

	Number				Size					
	RT		Error		RT		Error		Overall Mean	
Task	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	RT	Error
Baseline	308	37	.05	.04	276	44	.04	.03	292	.04
Filtering	327	40	.06	.05	274	38	.03	.02	301	.04
Positive correlation	302	38	.02	.03	279	43	.02	.03	290	.02
Negative correlation	285	37	.02	.02	281	38	.03	.02	283	.02
Overall mean	310		.04		277		.03		293	.03

Table 8
Stroop Congruity: Mean Reaction Times (in Milliseconds) to Congruent and Incongruent Stimuli for Judgments of Numerical and Physical Size in Baseline and Filtering Tasks (Experiment 4)

Task	Congruent		Incongruent		Congruity Score
	M	*SD*	*M*	*SD*	
Numerical size					
Baseline	315	36	302	39	−13
Filtering	334	41	320	42	−14
Physical size					
Baseline	273	43	280	46	+7
Filtering	271	39	277	38	+6

an incongruent stimulus, yet, at baseline, a large 6 was identified 74.5 msec more rapidly than was a small 6. In filtering, the difference in favor of the incongruent large 6 was 46 msec, and it was 49.2 msec in the correlated tasks. In each of the former tasks, the usual congruity effect was obtained for the other numeral, 8, yet the large incongruity scores for 6 sufficed to produce the negative Stroop measures in Tables 7 and 8. (The fast responses to a large 6 as well as the better legibility of a small 8 compared to that of a small 6 conspired to make the difference in RT between large and small numerals greater for 6 than for 8.) In each task, a significant congruity × stimulus interaction appeared [$F(1,13) = 137.76$, $MS_e = 51{,}575.58$ at baseline; $F(1,13) = 58.51$, $MS_e = 16{,}961.17$ in filtering; and $F(1,13) = 14.12$, $MS_e = 6{,}911.1$ for the correlated tasks]. This pattern of responding to the numeral 6, along with the expected pattern of responding to the numeral 8 (i.e., faster RTs for the larger size), engendered a significant difference at baseline for numerical classifications (large stimuli were classified more rapidly than were small stimuli [$t(13) = 11.71$, $SE_{dm} = 5.164$]; no comparable difference appeared for baseline size judgments between the two numerals). However, regardless of the direction of the Stroop effects observed, what is most revealing about the data of Experiment 4 is that the Stroop effects (along with a sizable Garner interference) were confined to the less discriminable numerical dimension.

A body of literature (summarized in Tzelgov et al., 1992) indicates that, naturally, numerals larger than 5 are treated as numerically large. As a result, the semantic conflict that drives the Stroop effect was not strong in the present case, since both numbers used in Experiment 4 were greater than 5. That also explains the predilection of our subjects to treat a large 6 as a congruent stimulus. Independently, subjects might also have responded more quickly to larger stimuli. Nevertheless, the pattern of results of Experiment 4 fits well with the data of Experiments 1–3, and it reinforces our conclusion that numerical value and physical size comprise optionally separable dimensions, the processing of which is constrained by relative discriminability.

Discussion

The typical reverse Stroop effect obtained in Experiment 4: Physical size strongly affected number identification, but numbers had a much weaker effect on judgments of size. Again, the more discriminable dimension intruded on the less discriminable dimension more pervasively than vice versa. The results of Experiment 4 thus are a mirror image of those found in Experiment 2. In Experiment 2, numerical value was the more discriminable dimension, and it interfered with judgments of size; in Experiment 4, physical size was the more discriminable dimension, and it interfered with judgments of number. Both sets of data support our conclusion that relative discriminability is a dominant factor underlying Stroop (and Garner) effects.

The results of Experiment 4 (and, to a degree, those of Experiment 3) reproduced interactions reported in the literature between numerical and physical magnitude of numerals: Physical size affects judgments of numerical value (see, e.g., Besner & Coltheart, 1979; Foltz et al., 1984; Hatta, 1983; Henik & Tzelgov, 1982; Tzelgov et al., 1992; Vaid, 1985). An explanation of those results espoused by many of the cited authors attributes the interaction to a reversed Stroop effect operating on the dimensions involved. Another explanation for those results, and one that we endorse, is that Stroop-like asymmetries are not intrinsic to numerical dimensions. The asymmetries arise from an inequality of the relative discriminabilities of values on the tested dimensions. The latter were not addressed through much of the research on numerical perception.

GENERAL DISCUSSION

Four experiments were conducted to elucidate multidimensional processing in numerical perception. On half the trials, the subjects classified numerical magnitude and attempted to ignore variation in the physical magnitude of the numerals; on the other half, subjects classified physical magnitude while attempting to ignore variation in numerical magnitude. When (approximately) matched in discriminability, numerical and physical size appeared separable; when mismatched, asymmetric Stroop and Garner effects occurred, with the more discriminable dimension intruding on the less discriminable dimension. When the dimensions were correlated positively, performance was better than that at baseline, and a Stroop effect, defined as the difference in performance between positively and negatively correlated dimensions, appeared throughout (i.e., even when no Garner and Stroop effects were obtained in the orthogonal contexts). The results of this study entail implications for numerical cognition, dimensional interaction and selective attention, and the understanding of the Stroop phenomenon itself. In the remainder of this article, we discuss each of these issues.

Selective Attention to Numerical Dimensions

Two interpretations of the overall pattern of these results are possible. According to the first interpretation, in orthogonal contexts, the observers were able to ignore the irrelevant dimension when the dimensions were matched in discriminability or when the irrelevant dimension was less discriminable than the relevant dimension, but were

unable to ignore the irrelevant dimension when it varied more noticeably than did the relevant dimension. In redundant tasks, the observers often attended to the irrelevant dimensions, perhaps because of the gain entailed in probability summation. However, once numbers were attended to—willingly or unwillingly, and whether the context was orthogonal or correlated—they were processed semantically (qua numbers, not as shapes). The second interpretation, and one that we prefer, is based on Pomerantz's (1991) notion that perceivers can exercise discretion regarding which dimensions to attend to in a speeded-classification task. According to this interpretation, in baseline and filtering, our subjects chose to ignore the semantic attributes of the numerals and classified shapes when they were instructed to classify numbers. Attending to shapes disrupts the intrusion potential of irrelevant size because size shares no quality of correspondence with relevant shape. The elimination of congruity explains the absence of Garner and Stroop effects in orthogonal contexts. In correlated contexts, by contrast, our subjects did follow the nominal instructions and attended to the semantic attributes of the numerals. They thus capitalized on the correspondences afforded by the semantic processing of numerals. Therefore, to maximize performance, our subjects were treating numbers as mere shapes in orthogonal tasks, but they were treating numbers as symbols for magnitude in correlated tasks. A unique form of dimensional interaction—optionally separable dimensions—ensued. That form bears on the description of numerical processing mandated by the present results, to the examination of which we turn next.

It cannot be emphasized too strongly that in the second interpretation we do *not* claim that subjects are engaging in some kind of controlled or deliberate strategic considerations that determine how the stimuli are processed. We do not assume conscious control of the contextually induced processing.

Numerical Cognition

Regardless of the interpretation that proves superior, our results show that numerical and physical magnitude of numerals constitute separable dimensions. Interactions, when they appear, are not intrinsic to dimensions involving numbers. Rather, they arise when the discriminabilities are grossly mismatched. The outcome of the present research conflicts with analog accounts of numerical perception. If, as those models hold, numbers are converted into perceptual magnitudes, changing the physical magnitudes of the numerals should interfere with numerical processing. It did not. Strong versions of the analog model hold that the activation of magnitude is obligatory: Presented with an arabic numeral, subjects cannot repress the activation of an underlying magnitude code, even when that representation is detrimental to performance. The present results conflict with this view: Our subjects could withstand interference from random variation in size when they were classifying number, and they similarly withstood interference from numbers when they were classifying size.

The present results portray a more involved picture of numerical processing than has previously been suspected. Clearly, a magnitude or ratio–scale representation is not mandatory for numerical perception. Interactions between numerical and physical attributes are rooted in performance asymmetries resulting from mismatched discriminabilities, not in conflicting (magnitude) representations. Our results, thus, are inconsistent with the main thrust of an "automaticity hypothesis" (see, e.g., Besner, Smith, & MacLeod, 1990) for numerical perception in which mandatory processing of numerical value is assumed. If one espouses our semantic interpretation, further conclusions follow. On that account, when it was advantageous, the subjects could jettison the semantic component of the presented numerals—treat the numerals as mere graphic stimuli—the instructions to attend to numbers notwithstanding. Seen from that vantage, Stevens's (1951) statement that a number "sometimes means the physical ink mark on a piece of paper" (p. 22) is particularly instructive. Patently, an 8 is only greater than a 6 when the number stimuli are processed *symbolically*, when they *stand for* referent magnitudes or quantities. Many contexts—correlated dimensions, magnitude estimation of sensory events, comparative judgment of numbers—induce that mode of processing. However, in other contexts, numbers need not function psychologically as numbers at all—they need not be symbols for magnitude.

Hinrichs and Novick (1982) examined the retention of numbers described to subjects either as car prices (magnitude mode) or telephone numbers (nominal mode) and found evidence for multiple representations for number, commensurate with the results of this study. Similarly, neuropsychological research on acalculia or acquired dyscalculia (see, e.g., Corballis & Sergent, 1992; McCarthy & Warrington, 1990; McCloskey, 1992; Sergent, 1990; Shallice, 1988) has provided striking examples of dissociations between symbolic and nominal processing of numbers in accord with the present scheme. For example, McCloskey and Caramazza (1987) reported both patients who were impaired in naming individual arabic numerals but who nonetheless organized number strings in the correct order of magnitude and those who showed the reverse deficit. In addition, McCarthy and Warrington (1990) concluded that the failure to comprehend numbers and retrieve number words are separate deficits, both of which can be selectively spared or impaired in patients.

Dimensional Interaction and the Stroop Phenomenon

Informed by the present results, we may gain new insight into dimensional interaction and view afresh the Stroop phenomenon itself. An often overlooked chapter in Garner's (1974) seminal work is titled, "Information Integration with Separable Dimensions" (chap. 7). There, Garner largely anticipated our probability summation hypothesis, observing that, for errors, "two processes must be independent if together they are to improve per-

formance" (p. 163). Even more poignant is his conclusion: "Information integration *can* occur with separable redundant dimensions" (p. 175, emphasis added). Earlier, Garner and Morton (1969) had shown that perceptual independence is not a unitary concept, and that several definitions must be considered in empirical contexts. Ashby, Maddox, Townsend, and their colleagues (e.g., Ashby, 1992; Ashby & Maddox, 1990; Ashby & Townsend, 1986; Maddox, 1992; Townsend, Hu, & Ashby, 1980) have developed that theme into a comprehensive theory of perceptual independence, called the General Recognition Theory (GRT; see also Multidimensional Signal Detection Theory, e.g., Kadlec & Townsend, 1992; Melara, 1992). GRT distinguishes between perceptual and decisional independence, and, generally, between hidden perceptual responses and observable decisional responses. Maddox (1992; see also Ashby & Maddox, 1994) concluded that the Garnerian filtering task is a good test of perceptual separability. Gauged by that standard, the present dimensions are perceptually separable. Simultaneously, the substantial redundancy gains observed are suggestive of the failure of decisional separability. The present results are thus commensurate with fundamental findings within GRT as well as with Garner's intuitions, providing yet another example of separable dimensions that nonetheless interact under privileged contexts.

Finally, consider the implications of our results for a novel understanding of the original Stroop effect: Naming the color of the ink in which a word is printed is disrupted if the meaning of the word defines another color. Naming the word, by contrast, is not affected by conflicting print color. The belief that meaning is activated automatically just about whenever a word is processed for any purpose (Carr, 1992) has informed much of the pertinent theorizing. However, a substantial amount of evidence now suggests that the mere presentation of a word is not a sufficient condition for a full activation of its meaning (see, for instance, Carr, 1992, and references therein). Thus, Smith and Magee (1980) have suggested that, "a word can be named... even when the word's meaning is not known" (p. 375). Similarly, Glaser and Glaser (1989) conclude that a "semantic interpretation of the stimulus" does *not* apply "in the task of reading words aloud" (p. 40). According to the "lexical hypothesis" developed by Glaser (1992), "the word module provides no semantic processing, not even of abstract meaning or relations" (p. 67). Indeed, Glaser and Glaser (1989) failed to find the "semantic gradient" that has been previously suggested to govern Stroop (Klein, 1964). However, if the word in the typical Stroop stimulus is not (always) processed semantically, how does one explain the asymmetric interference that is usually (by no means invariably) reported?

The two interpretations of selective attention offered for the present results provide two novel theories of Stroop. The first (suggested by Melara & Mounts, 1993) attributes Stroop to a failure of selective attention caused by unequal discriminability. In an important study that invites reevaluation of the entire Stroop literature (estimated at 700 publications in 1991; see MacLeod, 1991), Melara and Mounts (1993) controlled the baseline discriminabilities of the original Stroop dimensions of word and color. Strikingly, when matched, colors and words appeared separable: Subjects were able to attend selectively to either words or colors without suffering interference from random variation on the irrelevant dimension. Both Stroop and Garner effects vanished when baseline discriminabilities matched. When mismatched, the more discriminable dimension disrupted classification on the less discriminable dimension. Melara and Mounts concluded that the Stroop phenomenon is an optional effect. Note that, in this interpretation, numbers (this study) or color words (Melara & Mounts, 1993) *are* processed semantically in all contexts.

The second interpretation offers an even more radical account of Stroop. By this theory, in orthogonal contexts, the subjects were classifying shapes when instructed to identify numbers (this study) or words (Melara & Mounts, 1993), wherefore neither stimulus was processed semantically. The theory also explains our failure to create the desired mismatch in favor of size in Experiment 3, because, presumably, at baseline, classification of shape was easier than classification of size. In general, then, in orthogonal contexts, subjects ignore the nominal instructions to attend to words and they classify forms instead. Semantic processing occurs only in correlated contexts, with the afforded correspondences facilitating performance. Therefore, in this theory, the Stroop effect is restricted to correlated contexts. One should recognize that the vast majority of extant Stroop studies (e.g., MacLeod, 1991) entail just such contexts by virtue of nonrandom allocation of colors to words and vice versa (note, incidentally, that Stroop's original study itself entailed a built-in correlation between word and color since he used no congruent stimuli).

In sum, in our context theory of the Stroop phenomenon, we claim that Stroop is an optional effect arising under either or both of the following contexts: Mismatched discriminability of colors and words (or values along any other Stroop-like dimensions) and nonrandom allocation of colors and words to the experimental stimuli. Conversely, matched discriminability and random allocation of dimensional values, as in orthogonal contexts, eliminate Stroop congruity or interference. We do not know yet whether both manipulations are necessary to eliminate the Stroop effects. Conceivably, failure to establish either context may result in the reinstatement of Stroop. Thus, when discriminability mismatched greatly, Stroop congruity appeared in several of the filtering conditions of both the Melara and Mounts and the present study. Future studies should pit discriminability against random allocation to firmly establish the necessary and sufficient conditions for the appearance of the Stroop effect.

We realize, of course, that our context theory of Stroop is fairly speculative at this point, and it is not the only interpretation mandated by our data. Nevertheless, context theory is compatible with much contemporary thinking

and research on Stroop (Lindsay & Jacoby, 1994; Melara & Mounts, 1993), multidimensional perception (Ashby, 1992; Ashby & Maddox, 1994), and semantic processing of nominally unattended stimuli (Carr, 1992). Careful scrutiny of our assumptions is necessary, of course, but the change in perspective may be important for a deep understanding of the Stroop effect and selective attention.

REFERENCES

ALGOM, D., & BABKOFF, H. (1984). Auditory temporal integration at threshold: Theories and some implications of current research. In W. D. Neff (Ed.), *Contributions to sensory physiology* (Vol. 8, pp. 131-159). New York: Academic Press.

ALGOM, D., DEKEL, A., & PANSKY, A. (1993). On the perception of number. In A. Garriga-Trillo, P. R. Minon, C. Garcia-Gallego, C. Lubin, J. M. Merino, & A. Villarino (Eds.), *Fechner Day '93* (pp. 1-6). Palma de Mallorca: International Society for Psychophysics.

ASHBY, F. G. (ED.) (1992). *Multidimensional models of perception and cognition*. Hillsdale, NJ: Erlbaum.

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

ASHBY, F. G., & MADDOX, W. T. (1994). A response time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology*, **38**, 423-466.

ASHBY, F. G., & TOWNSEND, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, **93**, 154-179.

BANKS, W. P., FUJII, M., & KAYRA-STUART, F. (1976). Semantic congruity effects in comparative judgments of magnitudes of digits. *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 435-447.

BESNER, D., & COLTHEART, M. (1979). Ideographic and alphabetic processing in skilled reading of English. *Neuropsychologia*, **17**, 467-472.

BESNER, D., GRIMSELL, D., & DAVIS, R. (1979). The mind's eye and the comparative judgment of number. *Neuropsychologia*, **17**, 373-380.

BESNER, D., SMITH, M. C., & MACLEOD, C. (1990). Visual word recognition: A dissociation of lexical and semantic processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 862-869.

BUCKLEY, P. B., & GILLMAN, C. B. (1974). Comparison of digits and dot patterns. *Journal of Experimental Psychology*, **103**, 1131-1136.

CARR, T. H. (1992). Automaticity and cognitive anatomy: Is word recognition "automatic?" *American Journal of Psychology*, **105**, 201-237.

CORBALLIS, M. C., & SERGENT, J. (1992). Judgments about numerosity by a commissurotomized subject. *Neuropsychologia*, **30**, 865-876.

DEHAENE, S. (1989). The psychophysics of numerical comparison: A reexamination of apparently incompatible data. *Perception & Psychophysics*, **45**, 557-566.

DEHAENE, S. (1992). Varieties of numerical abilities. *Cognition*, **44**, 1-42.

DEHAENE, S., DUPOUX, E., & MEHLER, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 626-641.

DEHAENE, S., & MEHLER, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, **43**, 1-29.

DUNCAN, E. M., & MCFARLAND, C. E., JR. (1980). Isolating the effects of symbolic distance and semantic congruity in comparative judgments: An additive-factors analysis. *Memory & Cognition*, **8**, 612-622.

FLOWERS, J. H., WARNER, J. L., & POLANSKY, M. L. (1979). Response and encoding factors in "ignoring" irrelevant information. *Memory & Cognition*, **7**, 86-94.

FOLTZ, G. S., POLTROCK, S. E., & POTTS, G. R. (1984). Mental comparisons of size and magnitude: Size congruity effects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 442-453.

FOX, L. A., SHOR, R. E., & STEINMAN, R. J. (1971). Semantic gradients and interference in naming color, spatial direction, and numerosity. *Journal of Experimental Psychology*, **91**, 59-65.

GALLISTEL, C. R., & GELMAN, R. (1992). Preverbal and verbal counting and computation. *Cognition*, **44**, 43-74.

GARNER, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.

GARNER, W. R. (1983). Asymmetric interactions of stimulus dimensions in perceptual information processing. In T. J. Tighe & B. E. Shepp (Eds.), *Perception, cognition, and development: Interactional analyses* (pp. 1-37). Hillsdale, NJ: Erlbaum.

GARNER, W. R., & FELFOLDY, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, **1**, 225-241.

GARNER, W. R., & MORTON, J. (1969). Perceptual independence: Definitions, models, and experimental paradigms. *Psychological Bulletin*, **72**, 233-259.

GARNER, W. R., PODGORNY, P., & FRASCA, E. M. (1982). Physical and cognitive dimensions in stimulus comparison. *Perception & Psychophysics*, **31**, 507-522.

GLASER, W. R. (1992). Picture naming. *Cognition*, **42**, 61-105.

GLASER, W. R., & GLASER, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General*, **118**, 13-42.

HATTA, T. (1983). Recognition of Japanese Kanji in the left and right visual field. *Neuropsychologia*, **21**, 669-678.

HENIK, A., & TZELGOV, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition*, **10**, 389-395.

HINRICHS, J. V., & NOVICK, L. R. (1982). Memory for numbers: Nominal vs. magnitude information. *Memory & Cognition*, **10**, 479-486.

HINRICHS, J. V., YURKO, D. S., & HU, J. M. (1981). Two-digit comparison: Use of place information. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 890-901.

HOCK, H., & PETRASEK, J. (1973). Verbal interference with perceptual classification: The effect of semantic structure. *Perception & Psychophysics*, **13**, 116-120.

JAFFE-KATZ, A., BUDESCU, D. V., & WALLSTEN, T. S. (1989).Timed magnitude comparisons of numerical and nonnumerical expressions of uncertainty. *Memory & Cognition*, **17**, 249-264.

KADLEC, H., & TOWNSEND, J. T. (1992). Signal detection analyses of dimensional interactions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 181-227). Hillsdale, NJ: Erlbaum.

KATZ, A. N. (1980). Cognitive arithmetic: Evidence for right hemispheric mediation in an elementary component stage. *Quarterly Journal of Experimental Psychology*, **32**, 69-84.

KLEIN, G. S. (1964). Semantic power measured through the interference of words with color-naming. *American Journal of Psychology*, **77**, 576-588.

LEFEVRE, J., BISANZ, J., & MRKONJIC, L. (1988). Cognitive arithmetic: Evidence for obligatory activation of arithmetic facts. *Memory & Cognition*, **16**, 45-53.

LINDSAY, D. S., & JACOBY, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 219-234.

LINK, S. W. (1990). Modelling imageless thought: The relative judgment theory of numerical comparisons. *Journal of Mathematical Psychology*, **34**, 2-41.

MACLEOD, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, **109**, 163-203.

MADDOX, W. T. (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 147-180). Hillsdale, NJ: Erlbaum.

MCCARTHY, A. M., & WARRINGTON, E. K. (1990). *Cognitive neuropsychology*. New York: Academic Press.

MCCLOSKEY, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, **44**, 107-157.

McCloskey, M., & Caramazza, A. (1987). Cognitive mechanisms in normal and impaired number processing. In G. Deloche & X. Seron (Eds.), *Mathematical disabilities: A cognitive neuropsychological perspective* (pp. 201-219). Hillsdale, NJ: Erlbaum.

Melara, R. D. (1989). Dimensional interaction between color and pitch. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 69-79.

Melara, R. D. (1992). The concept of similarity: From psychophysics to cognitive psychology. In D. Algom (Ed.), *Psychophysical approaches to cognition* (pp. 303-388). Amsterdam: Elsevier, North-Holland.

Melara, R. D., & Marks, L. E. (1990). Processes underlying dimensional interactions: Correspondences between linguistic and nonlinguistic dimensions. *Memory & Cognition*, **18**, 477-495.

Melara, R. D., & Mounts, J. R. W. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory & Cognition*, **21**, 627-645.

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'Brian, T. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, **116**, 323-336.

Morton, J. (1969). Categories of interference: Verbal mediation and conflict in card sorting. *British Journal of Psychology*, **60**, 329-346.

Moyer, R. S., & Landauer, T. K. (1967). The time required for judgments of numerical inequality. *Nature*, **215**, 1519-1520.

Parkman, J. M. (1971). Temporal aspects of digit and letter inequality judgments. *Journal of Experimental Psychology*, **91**, 191-205.

Poltrock, S. E., & Schwartz, D. R. (1984). Comparative judgments of multidigit numbers. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 32-45.

Pomerantz, J. R. (1983). Global and local precedence: Selective attention in form and motion perception. *Journal of Experimental Psychology: General*, **112**, 515-540.

Pomerantz, J. R. (1986). Visual form perception: An overview. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines: Visual perception* (pp. 1-30). New York: Academic Press.

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: Essays in honor of Wendell R. Garner* (pp. 195-210). Washington DC: American Psychological Association.

Pomerantz, J. R., Pristach, E. A., & Carson, C. E. (1989). Attention and object perception. In B. E. Shepp & S. Ballesteros (Eds.), *Object perception: Structure and process* (pp. 53-89). Hillsdale, NJ: Erlbaum.

Reisberg, D., Baron, J., & Kemler, D. G. (1980). Overcoming Stroop interference: The effects of practice on distractor potency. *Journal of Experimental Psychology: Human Perception & Performance*, **6**, 140-150.

Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology*, **83**, 274-278.

Sekuler, R., & Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Development*, **48**, 630-633.

Sekuler, R., Rubin, E., & Armstrong, R. (1971). Processing numerical information: A choice time analysis. *Journal of Experimental Psychology*, **90**, 75-80.

Sergent, J. (1990). Furtive incursions into bicameral minds. Integrative and coordinating role of subcortical structures. *Brain*, **113**, 537-568.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.

Shepard, R. N., Kilpatric, D. W., & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive Psychology*, **7**, 82-138.

Shor, R. E. (1971). Symbol processing speed differences and symbol interference effects in a variety of concept domains. *Journal of General Psychology*, **85**, 187-205.

Smith, M. C., & Magee, L. E. (1980). Tracing the time course of picture-word processing. *Journal of Experimental Psychology: General*, **109**, 373-392.

Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1-49). New York: Wiley.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, **18**, 643-662.

Sudevan, P., & Taylor, D. A. (1987). The cuing and priming of cognitive operations. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 89-103.

Takahashi, A., & Green, D. (1983). Numerical judgments with Kanji and Kana. *Neuropsychologia*, **21**, 259-263.

Townsend, J. T., Hu, G. G., & Ashby, F. G. (1980). A test of visual sampling independence with orthogonal straight lines. *Bulletin of the Psychonomic Society*, **15**, 163-166.

Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 166-179.

Vaid, J. (1985). Numerical size comparisons in a phonologically transparent script. *Perception & Psychophysics*, **37**, 592-595.

Vaid, J., & Corina, D. (1989). Visual field asymmetries in numerical size comparisons of digits, words, and signs. *Brain & Language*, **36**, 117-126.

Windes, J. D. (1968). Reaction time for numerical coding and naming of numerals. *Journal of Experimental Psychology*, **78**, 318-322.

Zbrodoff, N. J., & Logan, G. D. (1986). On the autonomy of mental processes: A case study of arithmetics. *Journal of Experimental Psychology: General*, **115**, 118-130.

(Manuscript received April 5, 1995;
revision accepted for publication August 21, 1995.)