Journal of Experimental Psychology:
Human Perception and Performance
1999, Vol. 25, No. 1, 39–58

0096-1523/99/$3.00

Stroop and Garner Effects in Comparative Judgment of Numerals: The Role of Attention

Ainat Pansky
Bar-Ilan University

Daniel Algom
Tel Aviv University

In 7 experiments, participants selected the larger member of pairs of digits that differed in numerical magnitude as well as in physical size. Selective attention to the relevant dimension (number or size) was gauged by Garner and Stroop interference, both of which varied considerably as a function of the number and relative discriminability of values along the constituent dimensions. When the to-be-ignored dimension was more discriminable, sizable Garner and Stroop effects affected performance on the relevant dimension. When it was less discriminable, Garner and Stroop effects were considerably smaller regardless of whether the relevant dimension was numerical or physical size. The sensitivity of Stroop interference to manipulations of discriminability is accounted for by the allocation of attention to the constituent dimensions. The demonstrated malleability of the Stroop effect is incompatible with claims of strong automaticity in numerical processing.

Our lives are informed by numbers at every turn. Numbers mark our age and intelligence, portray the balance in our checking accounts, and depict our height, weight, or blood pressure. The ubiquitous presence of numbers in our daily experience is easily accounted for by their role as symbols representing magnitude or quantity. In the latter realm, numbers comprise our words in the strictest linguistic sense (cf. Hurford, 1987). In some instances, though, a number may merely comprise a physical stimulus (as may, incidentally, a word). As Stevens (1951) pointed out, at times a numeral is simply an ink mark on a piece of paper. To say that 7 is larger than 3 is thus meaningful only if the ideographic signs function as symbols for referent magnitude. As mere graphic stimuli, incidentally, 7 is neither larger nor smaller than 3 (again, as mere physical stimuli, the word *cat* is neither larger nor smaller than the word *dog*). Therefore, to perform quantitative comparisons between numbers, people must retrieve the referent quantities or magnitudes. That we do so effortlessly and apparently without awareness is among the facts sustaining the widely accepted notion of automatic activation of semantic (magnitude) information for number.

According to the automaticity account, meaning—information related to numerical magnitude—is activated mandatorily just about whenever numerals are presented for any purpose. Versions of strong automaticity furthermore entail that the activation of magnitude is completely immune to task demands and attention, such that the strength of the semantic activation remains the same throughout attentional and task manipulations. Thus, in accord with the notion of automaticity (weak or strong), when people compare the physical size of numerals (i.e., a nonsemantic dimension of numerals), they usually perform better with congruent pairs (e.g., 2–8) than with incongruent pairs (e.g., 2–8), producing an appreciable Stroop effect. Presumably, performance is also superior for pairs that do not vary on the irrelevant dimension of numerical magnitude (e.g., 2–2) than for pairs that do (e.g., 2–8, 2–8), producing an appreciable Garner effect.

Are semantic (magnitude) representations called up ineluctably in numerical processing? If they are, can the strength of their activation be nonetheless brought under strict experimental control? In the current study, we sought to answer these questions by manipulating the saliency of the semantic and the physical dimensions. For the most part, we made values on one or the other dimension more discriminable and hence of greater priority in the allocation of attention. We used selective attention—gauged by Stroop and Garner interference—as our analytical tool. Our results do not accord with claims of strong automaticity for numerical perception. Although irrelevant semantic processing was activated in most (but not all) cases, its magnitude was systematically controlled by the allocation of attention. The more salient dimension, whether number *or* size, intruded on performance on the less salient dimension more than vice versa.

Stroop Effects in Numerical Processing

Studies of numerical comparison (see Algom, Dekel, & Pansky, 1996, and Dehaene, 1992, for recent reviews)

Ainat Pansky, Department of Psychology, Bar-Ilan University, Ramat-Gan, Israel; Daniel Algom, Department of Psychology, Tel Aviv University, Tel Aviv, Israel. Ainat Pansky is now at the Department of Psychology, University of Haifa, Haifa, Israel.

This research was conducted by Ainat Pansky under the supervision of Daniel Algom as part of the requirements for the doctoral degree at Bar-Ilan University. The research was supported by Israeli Trustees Foundation Grant DC-7/17 and by Israel National Science Foundation Grant 93-58. We thank Robert Melara and Todd Maddox for helpful comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Daniel Algom, Department of Psychology, Tel Aviv University, Ramat Aviv 69978, Tel Aviv, Israel. Electronic mail may be sent to algomd@freud.tau.ac.il.

largely used the design of Moyer and Landauer's (1967) pioneer study: Pairs of numerical stimuli differed solely in numerical value, and, as a result, the participants performed semantic comparisons of numerical magnitude. For these comparisons, Moyer and Landauer found a *symbolic distance effect* by which the larger the difference in numerical value between the digits, the shorter the time needed to decide which is the larger (see Banks, Fujii, & Kayra-Stuart, 1976; Buckley & Gillman, 1974; Garner, Podgorny, & Frasca, 1982, for a sample of the numerous reproductions of the effect). Note, however, that the same stimuli allow of physical comparisons if they vary along additional dimensions, such as the size of the presented numerals or their numerosity. The multidimensional avenue was pursued by Besner and Coltheart (1979), who presented participants with pairs of digits that differed in physical size as well as in numerical magnitude. On a trial, the relation between physical and semantic size could be congruent (as in the pair 2–8), incongruent (as in the pair 2–8), or neutral (as in the pair 2–8). The participants were instructed to choose the numerically larger member of a pair and to ignore physical size. Selective attention to numerical value failed, however, because responding was rapid for congruent, slow for incongruent, and intermediate for neutral pairs of stimuli. Irrelevant physical size thus intruded on judgments of numerical magnitude.

The interference found in Besner and Coltheart's (1979) study is related to the well-known interference of color words with color naming discovered by Stroop (1935). Note, though, that Besner and Coltheart only tested performance on the semantic dimension of numerical value (the equivalent of word reading in the standard Stroop task) and found intrusions from irrelevant physical size—an interference opposite in direction to that depicted by the classic Stroop effect.

Intrusions of physical size (or numerosity) on numerical value—reverse Stroop effects—have since been documented in several studies (e.g., Flowers, Warner, & Polansky, 1979; Foltz, Poltrock, & Potts, 1984; Hatta, 1983; Henik & Tzelgov, 1982; Morton, 1969; Shor, 1971; Takahashi & Green, 1983; Tzelgov, Meyer, & Henik, 1992; Vaid, 1985; Vaid & Corina, 1989; Windes, 1968). The classic Stroop effect, whereby judgments of physical size (or numerosity) are hampered by irrelevant numerical magnitude, has also been reported (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; Washburn, 1994; Windes, 1968; see also Dehaene & Akhavein, 1995, and Dehaene, Bossini, & Giraux, 1993, who found that irrelevant numerical magnitude interferes with same–different and parity judgments).

Are Stroop Effects Mandatory in Numerical Processing?

Stroop effects are often invoked as the prototype of automatic processing of the meaning of words and numbers. They demonstrate the involuntary activation of semantic information, because people appear to engage the meaning of words and numbers—to read the stimuli—in the face of instructions calling upon them to ignore meaning. The case for a mandatory Stroop effect in numerical processing seems especially strong. Obligatory activation of meaning is accomplished in this domain by assuming, in accord with the original idea of an "internal psychophysics" (Moyer & Landauer, 1967), that numbers are mapped onto an underlying magnitude code, the same magnitude code that sustains the processing of physical stimuli (e.g., Buckley & Gillman, 1974; Dehaene, Dupoux, & Mehler, 1990; Dehaene & Mehler, 1992; Gallistel & Gelman, 1992; Link, 1990; Restle, 1970; Sekuler, Rubin, & Armstrong, 1971; Shepard, Kilpatric, & Cunningham, 1975). This internal magnitude code is activated automatically upon presentation of the numerical stimulus (Dehaene, 1992; Dehaene & Akhavein, 1995; Dehaene et al., 1993; Sudevan & Taylor, 1987). Stroop interference ensues as a result of conflicting values of internal magnitude; one magnitude code represents numerical value, the other, physical size.

Automatic processes are fast, involuntary, ballistic, and do not draw on general resources (see, e.g., Hasher & Zacks, 1979, and Posner, 1978, for a discussion of these and further criteria for automaticity). The assumption of automaticity, in tandem with the relative speed of processing of the constituent dimensions, is used to explain the Stroop asymmetry: Because automatic processes are very fast and obligatory, color naming suffers intrusions from involuntary word reading, but not vice versa. Nevertheless, Kahneman and Treisman (1984; see also Kahneman & Chajczyk, 1983) distinguished between levels of automaticity. Processes are strongly automatic if they are completely immune to the effects of attention; they are partially automatic if they are completed without attention but are nonetheless facilitated by attention. Logan (1988) referred to theories positing strong automaticity as modal views of automaticity and explained their sundry implications.

In the current study, we manipulated the saliency of the semantic (numerical magnitude) and nonsemantic (physical size) dimensions of numerals by altering the discriminability of values along each dimension. We hypothesized that relative discriminability would determine the extent of attention allocated to the constituent dimensions, whether semantic or physical. In particular, if physical size is made more discriminable than numerical magnitude, then the speed and accuracy with which numerals are compared by size are only slightly affected by (irrelevant) numerical magnitude. Conversely, under the same pattern of discriminability, comparisons of numerical magnitude are affected much more by physical size, and the cost incurred to performance is expressed by sizable Stroop and Garner effects. Comparing numbers (or reading words) is considered to be an automatic process, yet we do expect the speed and accuracy with which numbers are compared (or words are read) to be affected by manipulation of discriminability. Moreover, we expect numerical comparisons to draw on general resources, namely, on the volume of attention directed to the relevant and irrelevant dimensions. Therefore, in contrast to others' versions of strong automaticity,

we predicted that the processing of numerals would be sensitive to variations in task and attentional demands.

In the current study we show that, in contrast to much extant research, the perception of number is not automatic in the strong sense. We propose instead an account of numerical perception based on the notion of *selective attention* to numerical dimensions. We show that selective attention is controllable experimentally and is subject to strategic influences. We do not claim, though, that these influences are necessarily conscious. Our results and conclusions challenge the traditional hypothesis of strongly automatic activation of meaning for alphanumeric stimuli. Our results are more compatible with a partially automatic view of numerical perception. We show that the Stroop effect is malleable by heretofore neglected experimental factors. By judicious manipulation of these factors, we were able to fabricate both the Stroop and reverse Stroop effects and to govern their magnitude in a lawful fashion.

In the remainder of this introduction, we discuss methodological issues plaguing extant research, provide evidence for the malleability of the Stroop effect for both word and number, distinguish between Stroop and Garner measures of selective attention, and present the rationale for the present set of experiments.

Discriminability of Number and of Physical Size of Numerals: A Neglected Factor in Numerical Stroop Research

The aforementioned study by Besner and Coltheart (1979) also illustrates the flaws and pitfalls associated with various Stroop investigations of numerical cognition. In their study, 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. Moreover, the pair of physical values was chosen in a completely arbitrary fashion. In a later study, Henik and Tzelgov (1982) had their participants judge physical size as well as numerical value. However, they also used seven values for numerical size (all digits between 2 and 8, inclusive), but only three values (heights of 4, 5, and 6 mm) for physical size. Again, no rationale was offered for the selection of values for physical size. The same semantic–physical asymmetry characterized the study of Tzelgov et al. (1992), and this seems typical. Virtually all pertinent research pitted a finely grained numerical dimension against a coarse physical dimension.

The bias in experimental design is important, because the asymmetry between the semantic (numerical) and physical dimensions 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. Moreover, in none of the reviewed experiments was there an attempt to assess, nay match, the *discriminability* of the numerical and physical dimension.

Discriminability specifies the psychological difference separating stimulus values along a dimension (Garner, 1983; Garner & Felfoldy, 1970; Melara & Mounts, 1993). It is an intradimensional index, measured separately for numerical value and physical size. Discriminability is matched if the values along the numerical dimension are as easily discriminable from one another as are the values along the dimension of physical size. Obviously, the unequal number of values used in earlier research precluded the possibility of an overall match across the physical and numerical magnitudes; neither is there any indication of the discriminability of the pairs of physical sizes that were used. However, as the following review shows, information on discriminability is critical for determining the magnitude and direction of Stroop effects in numerical processing.

The Malleability of the Stroop Effect for Word and Number

In a recent study, Melara and Mounts (1993) showed that Stroop interference is malleable, with the more discriminable dimension causing a failure of selective attention to the less discriminable dimension, but not vice versa. By varying the relative discriminability of the colors and words used, Melara and Mounts were able to fabricate both Stroop (when words were more discriminable than colors) and reverse Stroop (when colors were more discriminable than words) effects.

The crucial role of matched discriminability in numerical perception was demonstrated by Algom et al. (1996). We used Garner's (1974) speeded classification paradigm for the dimensions of numerical magnitude and physical size of numerals. A single number was shown on each trial, and the participant's task was to classify it on the relevant dimension—semantic or physical—as quickly as possible. The numbers also differed along the irrelevant dimension that the participant was instructed to ignore. In that study, we compared performance (reaction time [RT] and accuracy) across the following two conditions.

In the *filtering condition,* the participant was asked to classify values on one dimension—say, whether numbers were 3s or 7s—while ignoring irrelevant variation on the second dimension—say, whether the physical size of the number was large or small. In the *baseline condition,* the participant was again asked to classify values on the criterial dimension (e.g., whether numbers were 3s or 7s), but the values on the irrelevant dimension were held constant (e.g., all numbers were physically large). The ability to attend selectively was measured by comparing performance at baseline, where the irrelevant dimension was held constant, with performance in the filtering condition, where the two dimensions were varied orthogonally. Deficit in the latter, *Garner interference* (Pomerantz, 1986), signals the failure of selective attention. Dimensions characterized by significant amounts of Garner interference are called *interacting dimensions;* dimensions that do not produce appreciable amounts of Garner interference are labeled *separable dimensions.* The parity of performance at baseline and filtering marks perfect selectivity of attention (cf. Garner, 1974).

We also measured Stroop effects (in each task) by calculating the difference in classification RT between trials in which numerical value and physical size corresponded and trials in which they conflicted. Notably, we measured Garner and Stroop effects under conditions in which discriminabilities were matched, mismatched in favor of numbers, or mismatched in favor of size.

When discriminabilities were matched, numerical magnitude and physical size appeared separable: Participants were able to attend selectively either to number or to size 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 of the less discriminable dimension, producing Stroop effects (when numerical value was more discriminable) and reverse Stroop effects (when physical size was more discriminable). Algom, Dekel, and Pansky (1993) collected similar data for other numerical stimuli, and Melara and Mounts (1993) reported the same pattern of results for the original Stroop dimensions of color and word. Therefore, both Melara and Mounts and Algom et al. (1993, 1996) concluded that the Stroop phenomenon is an optional effect, reflecting the failure of selective attention caused by unequal discriminability.

These results bear strongly on earlier research, which has largely ignored the issue of discriminability. It is conceivable that the Stroop effects obtained in these studies were caused in part by an asymmetry in the baseline discriminability of the tested dimensions. If so, magnitude representations may not be mandatory for numerical perception. Even if they are mandatory, their magnitude may largely be dependent on relative discriminability.

Partitioning the Interference From Task-Irrelevant Sources of Information

The current study is the first to apply the Garner paradigm to pairwise comparisons between stimuli (rather than to classification of single stimuli, as in the original design). Application of the Garnerian logic to comparative judgments allows for a methodical partition of the observed interference. When neutral, congruent, and incongruent pairs of numerals are presented within the same block of trials (traditional research), interference entails a confluence of the following sources of irrelevant information: (a) Conflicting values of the relevant and irrelevant dimensions within a pair (e.g., 2–8; Stroop interference); (b) pair-to-pair changes in dimensional relations, with congruent, incongruent, and neutral pairs appearing in a random sequence (e.g., 2–8, 2–8, 2–2; a variety of Garner interference); and (c) pair-to-pair changes in the stimuli presented, a variation deriving from the use of multiple values of number and size (another variety of Garner interference). Presenting the neutral pairs in one block and congruent and incongruent pairs in a separate block (the hallmark of Garner's approach) eliminates the second source of irrelevant variation. Restricting the stimulus array to just a single pair of numerals and values of size (in a close approximation of the original Garnerian design) largely eliminates the third source. In Experiment 1, we controlled for the second and third sources of irrelevant interference, and in Experiment 2 we controlled for none. Examining all three design options in this study, we concluded that the one allocating the neutral and the other pairs into separate blocks (thus controlling for the second source of interference) along with presenting multiple values for number and size was best suited to elucidate the theoretical issues at hand (Experiments 3–7). We used this analysis of variance (ANOVA) model to elucidate the results of the individual experiments as well as the overall pattern of data of the current study.

The Present Experiments

This research was designed to elucidate the roles of automaticity and attention in the processing of semantic information in numerical perception. Earlier studies have found numerical value and physical size to interact, whereas Algom et al. (1993, 1996) showed numerical value and physical size to be separable dimensions. The former studies ignored discriminability of values along the tested dimensions; the latter measured and matched it. Nevertheless, we issue several caveats before drawing too firm a conclusion. First, previous studies have had participants compare pairs of numerals, whereas Algom et al. (1996) had participants classify single numerals. Note, incidentally, that the original Stroop design also entailed presentations of single stimuli (e.g., the word *green* printed in red), as did Garner's speeded classification paradigm. Another concern relates to the fairly contrived nature of the traditional Garner classification task. This task entails a small stimulus ensemble: The participant classifies two values along a dimension as the values alternate randomly from trial to trial. Although nothing in the paradigm requires that stimulus dimensions be binary valued, nonetheless both our studies and the great bulk of dimensional research (e.g., Ashby & Maddox, 1994; Garner, 1974; Maddox & Ashby, 1996; Melara, 1992) entailed testing only two values on each of two dimensions. Numerical Stroop research, by contrast, has typically used many numbers but much fewer values of physical size.

In the current study we therefore used comparative judgments of pairs of numerals. Moreover, we included many values of numerical magnitude and as many values of physical size. We also measured relative discriminability. Extensive pilot research permitted us to monitor and manipulate the relative discriminability of numerical value and physical size. Thus, we closely mimicked the traditional stimulus ensemble yet still controlled discriminability. Indeed, the hallmark of our approach was the rigorous control of discriminability along the tested dimensions. Emergence of uniformly large Stroop effects in the face of variation in discriminability would lend powerful support for the theory positing strong automatic activation of semantic information. Conversely, if Stroop interference is greatly affected by variations in relative discriminability, then the results pose a serious challenge to strongly automatic accounts of numerical cognition.

Experiment 1

The purpose of the first experiment was to extend our previous design and findings (Algom et al., 1993, 1996) to comparative judgments of numerals—the method used in the great bulk of traditional research. In the current experiment, the participants made comparative judgments, yet those judgments were made within the framework of a design that closely followed Garner's original paradigm. Our adaptation permitted us to gauge both Garner and Stroop interference for paired comparisons. The dimensions of numerical magnitude and physical size each assumed two values. Experimental stimuli were created from all possible combinations of these values. Most important, the experiment was preceded by extensive preliminary testing to match the baseline discriminability for numerical value and physical size.

Method

Participants. The participants were 20 Bar-Ilan University undergraduates who were paid to participate in the experiment. Their ages ranged from 19 to 25 years.

Stimuli and apparatus. Each member of the pair 2–8 appeared in each physical size: 18 × 12 pixels (small) and 28 × 19 pixels (large); a pixel extended 0.4 mm in any orientation. The smaller stimulus subtended 0.52° of visual angle in length and 0.34° in width. The respective visual angles for the larger stimulus were 0.8° and 0.54°. These values for number and size served to form congruent (2–8, 8–2), incongruent (2–8, 8–2), and unidimensional or neutral stimuli (2–8, 8–2, 2–8, and 8–2 for judgments of numerical value; 2–2, 2–2, 8–8, and 8–8 for judgments of physical size). From these stimulus pairs, we created four experimental tasks (two involving judgments of numerical magnitude and two involving judgments of physical size). For comparisons of numerical magnitude, participants performed a baseline task (in two blocks, with physical size held constant at either the large or the small value) and a filtering task (physical size varied orthogonally, with the pairs either congruent or incongruent). The participants also performed in two complementary tasks comparing physical size. A brief description of the tasks follows.

In the baseline tasks, the participants compared values on one dimension (e.g., numerical magnitude) while the other dimension was held at a constant value (e.g., both numerals were of small physical size). In other words, only neutral pairs were presented in these tasks. The baseline tasks included two blocks, one for each constant value on the irrelevant dimension (small or large). Each block consisted of 20 trials: 10 trials for each of the two stimulus pairs presented. In the filtering tasks, the participants again compared values on the criterial dimension (e.g., numerical magnitude), but the stimuli also differed along the irrelevant dimension (physical size). Congruent and incongruent pairs of stimuli alternated randomly from trial to trial. Each filtering task consisted of 40 trials: 10 trials for each of the four stimulus pairs presented.

The values of physical size used were determined on the basis of pilot testing. Our purpose was to equate the speeds of the numerical and physical comparisons at baseline. The stimuli were generated in Pascal small font by an IBM-compatible (PC 386) microcomputer and displayed on a super-VGA 14-in. (35.56 cm) color monitor. The stimuli appeared white over a dark background at the centers of the left and right hemifields; to avoid adaptation, we introduced a trial-to-trial spatial uncertainty of up to 5 pixels around the target locations. The viewing distance was approximately 80 cm from the center of the screen so that the stimuli appeared at 4.65° of visual angle to the right and left of the fixation point.

Procedure. The participants were tested individually in a dimly lit room. Each individual performed the two numerical or two physical comparison tasks together as a set, with half of the participants first performing the numerical tasks and half first performing the physical tasks. Within each set, half of the participants first performed the baseline tasks, and half the filtering task. For the former tasks, block order was varied in a random fashion. Prior to performing a particular task, the participants performed the entire set of trials of that task as practice. Trials were presented randomly within each task, subject to the proviso that no more than three stimuli with the same correct response appear in a sequence. Intervals of approximately 2 min separated the various tasks.

The participants were instructed to attend to the relevant dimension and select the larger member of the pair of numerals on that dimension. They were encouraged to respond quickly but accurately. Comparative judgments were made by pressing either a right- or left-hand key on the computer keyboard, according to the location of the larger stimulus on the screen. The stimuli were response terminated. A new pair of stimuli was presented following a 0.5-s interval. Reaction time was measured in milliseconds with a software timer.

Data analysis. The same general procedure was used to analyze the experiments reported here. First, trials in which the participant's RT was greater by more than 3 standard deviations than her or his mean for that type of trial (congruent, incongruent, or neutral) were excluded from the analyses. Planned comparisons among pairs of conditions were performed using the Bonferroni correction (.05 criterion value). Two global ANOVAs were routinely performed. The first tested for Garner interference; the second for Stroop congruity. Each analysis was performed in quadruplicate: separately for comparisons of numerical magnitude and physical size and then, within each criterial dimension, separately for RT and error rate. For Garner interference, task (baseline, filtering) and spatial organization (larger stimulus on the left or on the right) served as within-subject factors, and order of dimensional judgment (number first, physical size first) and task order (baseline first, filtering first) served as between-subjects factors. For Stroop congruity, RTs and error rates in the filtering task were analyzed with pair type (congruent, incongruent) and spatial organization (larger stimulus on the left or on the right) as within-subject factors and order of dimensional judgment (comparisons of numerical magnitude first, comparisons of physical size first) and task order (baseline first, filtering first) as between-subjects factors. In Experiments 3–5, ordinal distance served as an additional within-subject factor in each of the ANOVAs.

In all the experiments, the correlation between speed (RT) and accuracy (error percent) was calculated separately for each criterial dimension. Of the dozen speed–accuracy correlations calculated in the current study, only three were statistically significant. Most were negative, which indicated that some participants (a small minority in each experiment) emphasized speed or accuracy at the expense of the other. Most participants, though, did not, and indeed the pattern of means for errors and RTs largely corresponded in all experiments. The occasional deviations from this pattern do not jeopardize our conclusions regarding the effects of attention in numerical processing. In this experiment, the correlation was $-.72$ ($p < .001$) for comparisons of numerical magnitude and $-.33$ ($p > .1$) for comparisons of physical size, suggesting a speed–accuracy trade-off for the former.

Results

Mean RTs and proportions of errors for the comparisons of number and size appear in Table 1. For RT, average baseline performance was 395 ms for numerical comparisons and 410 ms for physical comparisons, indicating comparable difficulty of the two tasks. The 15-ms difference in performance was not significant, $t(19) = 1.92, p > .05$, confirming our success at matching baseline discriminability of the two dimensions. Error rates were also comparable at baseline; the 0.73% advantage for physical comparisons constituted an insignificant difference, $t(19) = 1.21, p > .2$.

The magnitude of Garner interference was fairly small for comparisons of both numerical value and physical size (7 and 15 ms, respectively). Statistical analyses, though, showed the effect to be significant for the latter, $F(1, 16) = 5.58, p < .05$, $MSE = 4{,}440.95$, but insignificant for the former ($F < 1$). However, another analysis—a Task × Dimension ANOVA—also showed that Garner interference did not differ significantly across number and size ($F < 1$ for the interaction). Our participants thus attended well to numerical magnitude in the face of variation in irrelevant physical size, although they attended somewhat less well to physical size in the face of variation in irrelevant numerical value.

For comparisons of number, error rates were actually lower in the filtering task (0.78%) than at baseline (1.62%), but insignificantly so, $F(1, 16) = 2.20, p > .1$. For comparisons of physical size, the corresponding values were 0.89% and 1.50%, and the difference was similarly insignificant, $F(1, 16) = 1.23, p > .2$. Therefore, the error data demonstrated that there was good selective attention both to numerical value and physical size in the face of random variation of values along the other, irrelevant dimension.

The Stroop congruity scores for the dimensions tested are also shown in Table 1. Stroop congruity is defined as the difference in performance to congruent pairs (in which the numerically larger member of the pair was physically larger) and incongruent pairs (in which the numerically larger member of the pair was physically smaller) in the filtering task. For comparisons of numerical value, responding was faster for the former by 17 ms, revealing a significant reverse Stroop effect, $F(1, 16) = 14.16, p < .003, MSE = 5{,}731.65$. For comparisons of physical size, congruent pairs were faster than incongruent pairs by merely 9 ms, an insignificant difference, $F(1, 16) = 2.51, p > .1$. The absence of a Stroop effect for physical comparisons is a striking result, showing that numerical magnitude did not intrude on judgments of physical size.

A glance at Table 1 also shows that the error data entail no Stroop effects for either type of judgment. For numerical comparisons, congruent pairs were responded to more accurately than incongruent pairs by merely 0.06%. For physical comparisons, the gain reaped by matching dimensional values amounted to 0.50%; neither congruity effect was statistically significant ($F < 1$ for both numerical and physical comparisons). The absence of a Stroop effect is particularly striking for comparisons of physical size. For those comparisons, both RT and error show that the semantic dimension of numerical value did not interfere strongly with the processing of the nonsemantic dimension of size.

Discussion

To appreciate the full theoretical import of these data, consider the subset of results obtained with comparisons of physical size—a nonsemantic dimension of number. First, the participants compared the physical size of numerals as accurately when the numerals differed in irrelevant numerical magnitude (filtering) as when they assumed the same numerical value (baseline). No Garner interference was obtained for errors. For RT, although the participants responded faster in the latter condition, the Garner interference was relatively small as well. Second, the participants compared pairs of numerals in which the numerically larger stimulus was also the physically larger as quickly and accurately as they did those in which the two dimensions

Table 1
Mean Reaction Times (RTs; in Milliseconds) and Error Rates for Comparison of Numerical Magnitude and Physical Size Across Task (Garner Interference) and Pair Type (Stroop Congruity) in Experiment 1

Assessment of Garner and Stroop effects	Numerical magnitude				Physical size			
	RT		Error (%)		RT		Error (%)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Garner analyses								
Task								
Baseline	395	48	1.62	2.86	410	65	0.89	2.26
Filtering	402	53	0.78	2.75	425	59	1.50	3.93
Garner interference	7		−0.84		15*		0.61	
Stroop analyses								
Pairs								
Congruent	394	53	0.75	2.67	421	57	1.25	3.35
Incongruent	411	55	0.81	2.87	430	63	1.75	4.46
Stroop congruity	17**		0.06		9		0.50	

*$p < .05$. **$p < .01$.

conflicted. The virtually identical performance with congruent and incongruent pairs—expressed by the absence of Stroop interference for both RT and error—shows that irrelevant numerical magnitude did not intrude on comparisons of physical size. These results mandate the following conclusion: Magnitude information is *not* accessed automatically from Arabic numerals under all circumstances.

In general, in the current experiment the absolute amount of interference, whether of Garner or Stroop species, was small. The results imply that when baseline discriminabilities are matched, participants can process the dimensions of numerical and physical magnitude fairly independently. We concluded that under certain conditions, numerical attributes can be separable dimensions.

The challenge to the automaticity account posed by the current results must be qualified, however, because we used a single pair of numerals, whereas many other studies included multiple pairs of numerals. In addition, espousing the Garner paradigm dictated the presentation of neutral pairs in one block and congruent and incongruent pairs in another. In virtually all the other studies, all three stimuli types were presented in a single block. In the next experiment, we eliminated the differences in method—presenting multiple pairs in a single block—with a single notable exception: Discriminability of the numbers and physical sizes used was rigorously controlled.

Experiment 2

In Experiment 2, congruent, incongruent, and neutral pairs of stimuli were presented within a single block—the popular method probing numerical Stroop effects (e.g., Besner & Coltheart, 1979; Foltz et al., 1984; Hatta, 1983; Henik & Tzelgov, 1982; Takahashi & Green, 1983; Vaid, 1985; Vaid & Corina, 1989). To further comply with the popular paradigm, the pair of values per dimension used in Experiment 1 was expanded to include many values of both number and size. However, unlike in previous studies, in this experiment we included an equal number of values for numerical magnitude and physical size. Most important, we selected values on the two dimensions such that equal "psychological distances" separated a given pair of stimuli along the two constituent dimensions. Consider, for instance, the pair 3–7. We composed the stimuli such that the speed and accuracy in comparing $_{3\text{-}7}$ (or 3–7; size held constant in baseline numerical judgment) were approximately the same as those for comparing $_{3\text{-}3}$ (or $_{7\text{-}7}$; number held constant in baseline size judgment). Our success at matching discriminability should not perhaps come as a surprise, as investigators have reported similar distance effects for number (Buckley & Gillman, 1974; Foltz et al., 1984; Moyer & Landauer, 1967) and size (Algom & Pansky, 1993; Curtis, Paulos, & Rule, 1973; Johnson, 1939).

Method

Participants. The participants were 20 Bar-Ilan University undergraduates who were paid for their involvement in the experiment. Their ages ranged from 19 to 27 years. None had participated in Experiment 1.

Stimuli and apparatus. For number, we used the seven digits between 2 and 8. For physical size, we used the following seven values (in pixels): (a) 18 × 12 (smallest), (b) 19 × 13, (c) 21 × 14, (d) 23 × 15, (e) 25 × 16, (f) 27 × 18, and (g) 28 × 19 (largest). A number was associated with a physical size such that each number appeared in that size throughout the experiment. The values of number and size were determined such that in any given pair of stimuli, the two numbers were as discriminable as the two sizes. To best attain the match in discriminability across the two dimensions, we did not use all the possible pairs and omitted those with adjacent and next-to-adjacent values of number and size. The following 10 pairs were selected: The numbers 2 and 5 with physical sizes a and d; 2 and 6 with a and e; 2 and 7 with a and f; 2 and 8 with a and g; 3 and 6 with b and e; 3 and 7 with b and f; 3 and 8 with b and g; 4 and 7 with c and f; 4 and 8 with c and g; 5 and 8 with d and g. For a given quadruplet, all 8 combinations of congruent, incongruent, and neutral pairs described in Experiment 1 were formed. The resulting 80 pairs constituted a block of trials. The apparatus, design, and conditions of stimulus presentation were the same as those used in Experiment 1.

Procedure. Again, half the participants performed comparisons of numerical magnitude first, and half performed comparisons of size first. A 2-min break separated performance in the two tasks. The participant was instructed to select the larger member of the pair of stimuli according to either numerical or physical size, as quickly and as accurately as possible. For each criterial dimension, four blocks of trials were run, the first of which served as practice. The participant could rest for as long as needed between the blocks. Within each block, order of trials was random. An entire experimental session, consisting of 160 practice trials and 480 experimental trials, lasted about 40 min.

Results

The overall correlation between speed and accuracy was $-.62$ ($p < .005$) for comparisons of numerical magnitude and $-.38$ ($p > .1$) for comparisons of physical size, suggesting a trade-off between the two measures for numerical comparisons. Reaction time is presented as a function of criterial dimension and pair type in Figure 1A. Our success at matching the discriminability of the two dimensions may be assessed by comparing performance on the neutral trials—pairs of stimuli differing only on the criterial dimension. For these stimuli, average RTs were 477 and 489 ms for numerical and physical comparisons, respectively. The 12-ms difference was not significant, $t(19) = 1.08$, $p > 0.2$, indicating that our attempt to match discriminability was successful.

The most striking feature of the comparative judgment functions presented in Figure 1 is the large effect of congruity, $F(2, 36) = 91.13$, $p < .001$, $MSE = 107{,}106.77$. For numerical comparisons, average RT was 452 ms on congruent trials and 509 ms on incongruent trials, indicating a reverse Stroop effect of 57 ms. Performance on the neutral trials (477 ms) partitioned Stroop effects into facilitation (the difference in RT between congruent and neutral pairs) and interference (the complementary difference between neutral and incongruent pairs). Planned comparisons confirmed that both facilitation (25 ms) and interference (32 ms) effects were statistically significant ($p < .001$). For physical comparisons, mean RTs were 465 and 552 ms for congruent and incongruent trials, producing a Stroop effect of 87 ms. For

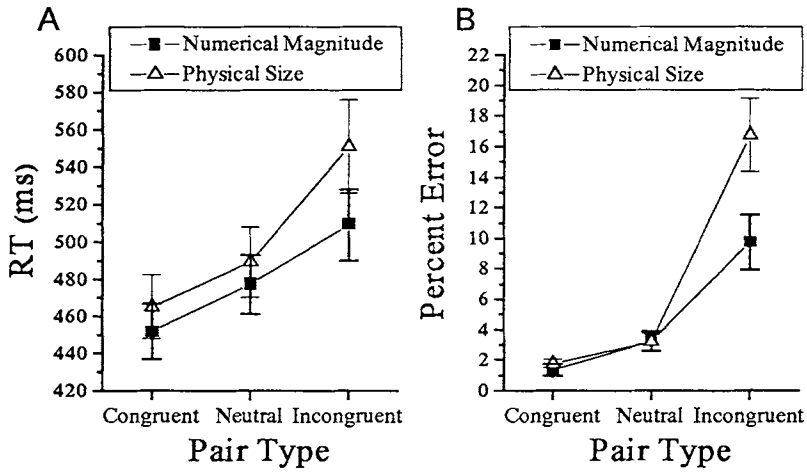

Figure 1. (A) Reaction time (RT) and (B) error rate as a function of criterial dimension and pair type in Experiment 2. The vertical bars mark one standard error around the mean.

these comparisons, interference (63 ms) was almost three times larger than facilitation (24 ms), although both effects were significant ($p < .001$).

For error rates (see Figure 1B), performance was again best on congruent trials, intermediate on neutral trials, and worst on incongruent trials: for congruity, $F(2, 36) = 37.27$, $p < .001$, $MSE = 2{,}229.78$. For numerical comparisons, error rates were 1.26%, 2.67%, and 8.27% for congruent, neutral, and incongruent pairs of stimuli, respectively. The respective data for physical size were 1.10%, 2.12%, and 13.48%. Partitioning the Stroop effect into interference and facilitation revealed statistically significant effects for both types of judgment ($p < .01$). Again, the nearly identical performance on neutral pairs (2.67% and 2.12%), $t(19) = 0.09$, $p > 0.9$, underscores our success at matching discriminability across the dimensions.

Discussion

Stroop effects (Henik & Tzelgov, 1982; Tzelgov et al., 1992) as well as reverse Stroop effects (Besner & Coltheart, 1979; Foltz et al., 1984; Henik & Tzelgov, 1982; Tzelgov et al., 1992; Vaid, 1985; Vaid & Corina, 1989) have been reported for comparisons of numerical and physical size. However, in many studies (e.g., Besner & Coltheart, 1979; Foltz et al., 1984; Vaid, 1985; Vaid & Corina, 1989), only numerical comparisons were examined, with physical size merely serving as a source of irrelevant information. As a result, these studies are moot on the role of attention and automaticity in numerical perception. Two studies (Henik & Tzelgov, 1982; Tzelgov et al., 1992) included both types of comparisons, but they did not match for discriminability. Because physical size was the more discriminable dimension in these experiments, it intruded on comparisons of numerical magnitude (reverse Stroop effect) more than did numerical magnitude on comparisons of physical size (Stroop effect). The imbalance in discriminability precludes a firm theoretical interpretation of these results as well.

We designed the current study to avoid dimensional imbalance by creating equal discriminability at baseline. We matched the discriminability of number and size by carefully selecting the values to create equal psychological distances along the two dimensions. Under the standardized conditions of this experiment, we detected sizable failures of selective attention for both number and size. Nevertheless, numerical value intruded more on judgments of size than did size on judgments of numerical value. Our results provide a critical set of data on numerical interactions, as our experimental conditions created no a priori bias favoring one dimension over the other.

The pattern of results obtained in Experiment 2 differs from that found in Experiment 1. In Experiment 1, selective attention to each dimension was good. In Experiment 2, by contrast, selective attention for both number and size was poor. The difference is readily explained by referring to our ANOVA model of interference. In Experiment 1, only two values per dimension were used with the neutral stimuli presented in a separate (baseline) block of trials. In Experiment 2, seven values per dimension were used with congruent, incongruent, and neutral stimuli presented within a single block of trials. Experiment 2 thus entailed two sources of irrelevant variation (the second and third sources in the model) missing in Experiment 1. As a result, selective attention failed in Experiment 2, expressed by the large values of Stroop interference observed. In the General Discussion section, we elucidate the mechanism producing the various types of interference.

In the next experiment, we used the method of Experiment 1: Neutral pairs were included in a baseline task, and congruent and incongruent pairs were included in a separate filtering task. This Garnerian paradigm provides better control of irrelevant variation (eliminating the second source in the ANOVA model) than does the current paradigm in which only a single block of presentations was used. However, in contrast to Experiment 1, in Experiment 3 we

included multiple pairs of stimuli (those used in this experiment). This extension of the Garnerian design enhances its ecological appeal and validity.

Experiment 3

Method

Participants. Twenty Bar-Ilan University undergraduates were paid to participate. Their ages ranged from 19 to 24 years. None had taken part in any of the previous experiments.

Stimuli and apparatus. The stimuli were those used in Experiment 2: all the combinations—congruent, incongruent, and neutral—created from the 10 pairs of numerals matched for discriminability of number and size. However, the method was that used in Experiment 1: Neutral pairs were presented in one block of trials (baseline task) and congruent and incongruent pairs in another block of trials (filtering task). For each criterial dimension, the baseline task contained the pairs in which the value of the irrelevant dimension was held constant. The filtering tasks contained the pairs in which the irrelevant dimension varied, with values either corresponding (congruent pairs) to or conflicting (incongruent pairs) with those of the relevant dimension.

Note that in this experiment, the baseline and filtering tasks included multiple pairs of stimuli. In Experiment 1, as well as in mainstream Garner research, a single pair of values per dimension was used. Therefore, in this experiment, at both baseline and filtering, there was considerable trial-to-trial variation in values of the to-be-judged dimension. In the filtering tasks, values on the irrelevant dimension changed as well, but they were kept constant within each pair in the baseline tasks. Apparatus, stimulus presentation, and viewing conditions were the same as those in the previous experiments.

Procedure. Each person participated in two (identical) experimental sessions, separated by at least 24 hr. In each session, half the participants compared numerical magnitude first, and half compared physical size first. For each criterial dimension, half the participants performed the baseline task first, and half performed the filtering task first. At the beginning of each task, the participants performed one block of 40 trials as practice. They then proceeded to perform three blocks of the 40 trials, resulting in a total of 120 judgments per task. The participants could rest for as long as needed between blocks. An entire experimental session, consisting of 160 practice trials and 480 experimental trials, lasted about 40 min.

Results

The correlation between speed and accuracy was .27 ($p > .2$) for comparisons of numerical magnitude and .30 ($p > .2$) for comparisons of physical size; no evidence of a trade-off between the two measures of performance was thus indicated. The main results are presented in Table 2. First, consider performance on the neutral pairs of the baseline task—stimuli differing solely on the relevant dimension. Average baseline RTs were 439 ms for numerical magnitude and 431 ms for physical size. The difference of 8 ms was not significant, $t(19) = 1.51, p > .1$, indicating that our attempt to match discriminability was as successful in this experiment as it was in Experiment 1. The insignificant difference of 0.19% in error rate, $t(19) = 0.62, p > .5$, provides additional evidence of matched discriminability at baseline.

Next consider the quality of selective attention obtained for number and size. For both types of comparisons, selective attention failed, as performance in filtering was worse than that at baseline. Values of Garner interference were 30 ms for number, $F(1, 16) = 43.82, p < .001, MSE = 135{,}560.08$, and 40 ms for size, $F(1, 16) = 47.70, p < .001, MSE = 244{,}612.34$. The two did not differ significantly, $F(1, 18) = 1.55, p > .2$. For error, Garner interference amounted to 0.87% for numerical comparisons, $F(1, 16) = 29.89, p < .001, MSE = 124.88$, and 1.97% for physical comparisons, $F(1, 16) = 21.75, p < .01, MSE = 120.10$, and again the two did not differ significantly, $F(1, 18) = 4.07, p > .05$. Therefore, random variation on an irrelevant dimension intruded on comparisons performed on the relevant dimension whether the participants were comparing numbers or sizes.

Analyses of Stroop congruity were confined to the

Table 2
Mean Reaction Times (RTs; in Milliseconds) and Error Rates for Comparison of Numerical Magnitude and Physical Size Across Task (Garner Interference) and Pair Type (Stroop Congruity) in Experiment 3

Assessment of Garner and Stroop effects	Numerical magnitude				Physical size			
	RT		Error (%)		RT		Error (%)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Garner analyses								
Task								
Baseline	439	60	2.15	4.55	431	55	2.34	4.41
Filtering	469	79	3.02	7.05	471	75	4.31	9.28
Garner interference	30**		0.87**		40**		1.97**	
Stroop analyses								
Pairs								
Congruent	439	62	0.44	2.11	438	56	0.65	3.45
Incongruent	499	82	5.60	9.05	503	78	7.98	11.56
Stroop congruity	60**		5.16**		65**		7.33**	

**$p < .01$.

filtering task, because only that task contained congruent and incongruent pairs of stimuli. As shown in Table 2, Stroop congruity was 60 ms for comparisons of numerical magnitude and 65 ms for comparisons of physical size; both effects were highly significant: for numerical comparisons, $F(1, 16) = 143.10$, $p < .001$, $MSE = 664{,}621.02$; for physical comparisons, $F(1, 16) = 157.72$, $p < .001$, $MSE = 673{,}876.43$. These effects did not differ significantly, $F(1, 18) = 1.37$, $p > .2$. For accuracy, Stroop congruity amounted to 5.16% for numerical comparisons, $F(1, 16) = 75.04$, $p < .001$, $MSE = 4{,}359.92$, and 7.33% for physical comparisons, $F(1, 16) = 97.50$, $p < .001$, $MSE = 8{,}585.12$. Again, the two did not differ significantly, $F(1, 18) = 3.68$, $p > .05$.

A surprising result of Experiment 3 was that congruent pairs (in the filtering task) were *not* compared faster than neutral pairs (in the baseline task). For physical size, congruent pairs were actually compared slightly slower (7 ms) than neutral pairs, although not significantly so, $t(19) = 1.41$, $p > .1$. For numerical comparisons, RT was identical for congruent and neutral pairs. In other words, the participants did not reap gain when the irrelevant dimension covaried with the criterial dimension, compared with a condition in which the irrelevant dimension was held constant. Our ANOVA model readily accommodated this result.[1,2]

Discussion

In this experiment, we selected numbers and physical sizes such that equal psychological distances separated corresponding values along the two dimensions. Under these standardized conditions, substantial Garner and Stroop effects emerged, signaling the failure of selective attention for either number or size. Moreover, the interference was comparable for judgments of number and size. The results corroborate earlier findings on the automatic activation of numerical information. That granted, our interpretation stresses the importance of attention in determining the magnitude of the respective amounts of interference. As the following two experiments show, the magnitude of interference is crucially dependent on the relative discriminability of the tested dimensions. If the to-be-ignored dimension is more discriminable than the to-be-attended one, it will intrude more on the latter, regardless of whether the participants are judging number or size.

The results of Experiment 3 differed in important respects from those of both the previous experiments. Our ANOVA model is singularly instructive in explicating the meaning of these differences. Experiments 1 and 3 differed only with respect to the size of the stimulus set used (the third source of irrelevant variation in our model): large in Experiment 3, small in Experiment 1. The presence in Experiment 3 of this additional source of irrelevant variation sufficed to engender substantial amounts of Garner and Stroop interference. As a result, in Experiment 1, the latter measures amounted to approximately one sixth of their value in Experiment 3. Experiments 2 and 3 differed with respect to another source of irrelevant variation (the second source in our model): Neutral pairs were presented along with congruent and incongruent pairs in a single block in Experiment 2, but they were presented in a separate block in Experiment 3. The trial-to-trial changes in dimensional relations impaired performance with the neutral pairs in Experiment 2. Performance with neutral pairs was free of this source of interference in Experiment 3. As a result, performance with neutral pairs improved in this experiment and was on a par with congruent pairs.

To test the strategic influences on selective attention to numerical dimensions, in the following two experiments we purposely manipulated relative discriminability such that one dimension was more discriminable than the other. A strong version of automaticity predicts comparative amounts of interference regardless of relative dimensional discriminability. By contrast, systematic changes in the pattern of interference would attest to the presence of considerable strategic control in numerical processing.

In Experiment 4, we reduced the differences in physical size, thereby making numerical magnitude more discriminable. In Experiment 5, we increased the differences in size, making physical size the more discriminable dimension. We emphasize that throughout the current experiments, all the values were completely discriminable psychophysically (i.e., any two adjacent values well exceeded the difference threshold). Our discriminability variable denotes psychological separation of dimensional values *beyond* psychophysical resolution (the minuscule rates of error obtained even for the less discriminable dimension show that both dimensions were fully discriminable in the psychophysical sense).

Experiment 4

Method

Participants. The participants were 20 Bar-Ilan University undergraduates who had not participated in any of the previous experiments. Their ages ranged from 19 to 26 years. They were paid for their participation.

Apparatus and stimuli. The numbers were the same seven digits used in Experiments 2 and 3. However, the physical sizes of

[1] However, we did find substantial interference, as incongruent pairs were compared more slowly than neutral pairs were. Interference was 60 ms for numerical comparisons, $t(19) = 10.19$, $p < .001$, and 72 ms for physical comparisons, $t(19) = 10.27$, $p < .001$, indicating that Stroop congruity consisted entirely of the toll exacted by conflicting values on the irrelevant dimension.

[2] For error rates, both facilitation and interference effects emerged. Error rates were smaller for congruent pairs than for neutral pairs by approximately 1.7% for both types of comparisons: for numerical comparisons, $t(19) = 6.19$, $p < .001$; for physical comparisons, $t(19) = 4.45$, $p < .001$, evincing redundancy gains for congruent pairs. Interference amounted to 3.45% for numerical comparisons, $t(19) = 8.30$, $p < .001$, and 5.64% for physical comparisons, $t(19) = 5.13$, $p < .001$. Hence, interference was twice as large as facilitation for numerical comparisons and three times as large as facilitation for physical comparisons. Overall, then, accuracy was best for congruent pairs, intermediate for neutral pairs, and worst for incongruent pairs.

the numerals were changed to make them closer and, hence, less discriminable than the numbers. The new values of physical size (in pixels) were (a) 18 × 12 (smallest), (b) 19 × 12, (c) 19 × 13, (d) 20 × 13, (e) 21 × 14, (f) 22 × 15, and (g) 23 × 15 (largest). The smallest stimulus subtended 0.8° and 0.54° of visual angle in length and in width, respectively, as in the previous experiments. The corresponding angles for the largest stimulus were 0.66° and 0.43°. As in Experiments 2 and 3, each number was associated with a physical size and was presented in that size throughout the experiment. Because physical size was more densely grained in this experiment, it was important to ensure that all values of size were discriminably different from one another (i.e., that any two neighboring values of size well exceeded the difference threshold). To make physical size fully discriminable psychophysically (based on pilot testing), we elected to include pairs separated by an ordinal distance of 4 or more on either dimension. The six pairs included were the numbers 2 and 6 with physical sizes a and e; 2 and 7 with a and f; 2 and 8 with a and g; 3 and 7 with b and f; 3 and 8 with b and g; and 4 and 8 with c and g. As before, for each quadruplet, all eight combinations of neutral, congruent, and incongruent pairs were created. As in Experiment 3, the baseline task for each criterial dimension consisted of the neutral pairs, and the filtering tasks consisted of the congruent and incongruent pairs. Apparatus, stimulus presentation, and viewing conditions were the same as in the previous experiments. Analyses of the data followed those of Experiment 3.

Procedure. The procedure was identical to that of Experiment 3, except that because of the smaller number of pairs used, only one experimental session was required. For each criterial dimension, seven blocks of trials were run, the first two of which served as practice. An entire experimental session consisting of 192 practice trials and 480 experimental trials lasted about 40 min.

Results

The correlation between speed and accuracy was −.13 ($p > .5$) for comparisons of numerical magnitude and −.41 ($p > .07$) for comparisons of physical size; no significant trade-off between the two measures of performance was thus indicated. Performance at baseline was one critical item to observe, because it showed whether number was, in fact, more discriminable than size. As shown in Table 3, average baseline RT was 451 ms for number and 541 ms for physical size. Therefore, our participants compared numerical magnitude with size held constant a full 90 ms faster than they did physical size with numerical value held constant, $t(19) = 7.65$, $p < .001$. The participants also made fewer errors in the former task (0.86%) than in the latter (2.10%). The difference of 1.24% was again significant, $t(19) = 2.83$, $p < .011$. Our manipulation was therefore successful in rendering number more discriminable than size.

Consider next selective attention to number and size, the main burden of the experiment. For both dimensions, the comparisons took longer in filtering than at baseline: for numerical comparisons, $F(1, 16) = 9.07$, $p < .01$, $MSE = 20{,}369.80$; for physical comparisons, $F(1, 16) = 28.03$, $p < .001$, $MSE = 243{,}838.97$. However, Garner interference was more than three times larger for size (64 ms) than for number (18 ms), and the difference was highly significant, $F(1, 18) = 16.22$, $p < .001$, $MSE = 63{,}518.65$. Therefore, relative discriminability engendered substantial changes in the pattern of numerical interaction. Comparisons of size were disrupted by irrelevant variation in numerical value much more than were comparisons of numerical value by irrelevant variation in size. The error data confirm to an even greater extent the crucial importance of relative discriminability. For comparison of number, error rates were 0.86% and 1.38% at baseline and in filtering, respectively, amounting to a minuscule Garner interference of 0.52%, which was not significant, $F(1, 16) = 1.42$, $p > .2$. For comparison of size—the less discriminable dimension—Garner interference was 1.64%, again more than three times that for number, $F(1, 16) = 6.52$, $p < .05$, $MSE = 162.51$.

The Stroop data, presented in Table 3, also show the large asymmetry in the quality of selective attention for the two dimensions. Congruent pairs were compared faster than

Table 3
Mean Reaction Times (RTs; in Milliseconds) and Error Rates for Comparison of Numerical Magnitude and Physical Size Across Task (Garner Interference) and Pair Type (Stroop Congruity) in Experiment 4

Assessment of Garner and Stroop effects	Numerical magnitude				Physical size			
	RT		Error (%)		RT		Error (%)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Garner analyses								
Task								
Baseline	451	53	0.86	2.48	541	84	2.10	4.31
Filtering	469	68	1.38	4.77	605	122	3.74	8.72
Garner interference	18**		0.52		64**		1.64*	
Stroop analyses								
Pairs								
Congruent	455	63	0.49	2.92	551	97	0.78	3.73
Incongruent	482	70	2.27	5.96	658	121	6.72	11.00
Stroop congruity	27**		1.78*		107**		5.94**	

*$p < .05$. **$p < .01$.

incongruent pairs for both number, $F(1, 16) = 42.51$, $p < .001$, $MSE = 43{,}250.13$, and size, $F(1, 16) = 78.22$, $p < .001$, $MSE = 691{,}813.90$, yet Stroop congruity was four times as large for size as for number, and the difference was highly significant, $F(1, 18) = 57.63$, $p < .001$, $MSE = 193{,}055.85$. For errors, too, a significant Stroop effect emerged both for number, $F(1, 16) = 5.33$, $p < .05$, and size, $F(1, 16) = 25.39$, $p < .001$, $MSE = 2{,}116.18$, but the latter effect was over three times larger than the former. Indeed, the interaction between dimension and Stroop congruity was highly significant, $F(1, 18) = 18.97$, $p < .001$, $MSE = 517.36$.

Discussion

In Experiment 4, we purposely selected values of physical size that were less discriminable than the corresponding values of numerical size. The manipulation of relative dimensional discriminability had a dramatic effect on the pattern of interaction between the two dimensions. Although Garner interference appeared for both dimensions, its magnitude was much larger for the physical comparisons of size than for the semantic comparisons of number. The asymmetrical pattern of interference was obtained for Stroop congruity as well. The Stroop effect obtained for comparisons of size was four times the value obtained for comparisons of number, although the latter also reached significance.

The results of Experiment 4 demonstrated the crucial importance of relative discriminability for the pattern of interaction observed between the dimensions of number and size. In Experiment 3, number and size were equally discriminable, and each suffered equivalent amounts of interference. In Experiment 4, in which number was more discriminable than size, the pattern of interference was asymmetrical, with size suffering more interference from number than number from size.

In Experiment 4, our participants were strongly influenced by numerical information even when irrelevant or detrimental to the task at hand. We attributed part of this strong semantic interference to the greater discriminability of numbers. According to the traditional version of strong automaticity, the semantic interference is expected to remain constant, regardless of attentional factors. To elucidate the different predictions, in Experiment 5 we selected values of physical size to be more discriminable than the values of numerical magnitude. We tested the extent to which irrelevant semantic information (numerical magnitude) interferes with comparisons of size when it constitutes the less discriminable dimension.

Experiment 5

Method

Participants. The participants were 20 Bar-Ilan University undergraduates who had not participated in any of the previous experiments. Their ages ranged from 20 to 25 years. They were paid to participate in the experiment.

Apparatus and stimuli. The numbers were the same seven digits used in the previous experiment. The physical sizes of the numerals were changed, however, to make them more discriminable than the numbers. The values of physical size (in pixels) were (a) 18 × 12 (smallest), (b) 31 × 21, (c) 45 × 30, (d) 58 × 39, (e) 72 × 48, (f) 85 × 57, and (g) 99 × 66 (largest). Again, the smallest stimulus subtended 0.8° and 0.54° of visual angle in length and in width, respectively. The corresponding angles for the largest stimulus were 2.83° and 1.89°. Each numeral was again associated with a given physical size and was displayed in that size throughout the experiment. Pilot testing showed that an advantage for physical size was obtained for pairs separated by an ordinal distance of 2 or more on both dimensions. The 15 pairs satisfying this criterion were the numbers 2 and 4 with physical sizes a and c; 2 and 5 with a and d; 2 and 6 with a and e; 2 and 7 with a and f; 2 and 8 with a and g; 3 and 5 with b and d; 3 and 6 with b and e; 3 and 7 with b and f; 3 and 8 with b and g; 4 and 6 with c and e; 4 and 7 with c and f; 4 and 8 with c and g; 5 and 7 with d and f; 5 and 8 with d and g; and 6 and 8 with e and g. For each quadruplet, all eight combinations of neutral, congruent, and incongruent pairs were created. The baseline task for each criterial dimension consisted of the neutral pairs; the filtering tasks consisted of the congruent and incongruent pairs. Apparatus, stimulus presentation, and viewing conditions were the same as those in the previous experiments.

Procedure. The procedure was identical to that of Experiment 3 except that, because of the larger number of pairs used, three blocks of trials were run for each criterial dimension, the first serving as practice. Each experimental session consisted of 240 practice trials and 480 experimental trials and lasted about 40 min.

Results

The correlation between speed and accuracy was $-.21$ ($p > .3$) for comparisons of numerical magnitude and $-.01$ ($p > .97$) for comparisons of physical size; no evidence of a trade-off between the two measures of performance was thus indicated. The baseline data, presented in Table 4, help to determine whether our manipulation was successful in rendering size more discriminable than number. The average time required to compare physical size at baseline (i.e., with number held constant; 359 ms) was 76 ms less than that needed to compare numerical magnitude (i.e., with size held constant; 435 ms), $t(19) = 10.36$, $p < .001$. The baseline error rate was also lower for size (1.29%) than for number (2.34%); the difference of 1.05% was again significant, $t(19) = 4.24$, $p < .001$. Therefore, both speed and accuracy showed the advantage of size over number in relative discriminability.

Garner interference was 51 ms for comparisons of number but only 12 ms for comparisons of size. Both effects were statistically significant: for comparisons of number, $F(1, 16) = 48.26$, $p < .001$, $MSE = 549{,}176.28$; for comparisons of size, $F(1, 16) = 6.80$, $p < .05$, $MSE = 549{,}176.28$. The difference in magnitude of Garner interference across the two dimensions was highly significant, $F(1, 18) = 28.08$, $p < .001$, $MSE = 80{,}164.38$. The error data mirrored the RT data. For physical comparisons, the small Garner interference of 0.36% was not significant, $F(1, 16) = 4.11$, $p > .05$. For comparisons of number, we found a larger Garner interference of 1.41%, which was statistically significant, $F(1, 16) = 12.47$, $p < .005$, $MSE = 395.16$.

Duplicating the pattern of Garner interference, Stroop congruity was significantly larger for number than for size,

Table 4
Mean Reaction Times (RTs; in Milliseconds) and Error Rates for Comparison of Numerical Magnitude and Physical Size Across Task (Garner Interference) and Pair Type (Stroop Congruity) in Experiment 5

Assessment of Garner and Stroop effects	Numerical magnitude				Physical size			
	RT		Error (%)		RT		Error (%)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Garner analyses								
Task								
Baseline	435	52	2.34	4.89	359	43	1.29	3.82
Filtering	486	86	3.75	8.80	371	61	1.65	5.13
Garner interference	51**		1.41**		12*		0.36	
Stroop analyses								
Pairs								
Congruent	448	72	1.41	4.66	359	48	0.99	4.09
Incongruent	525	82	6.09	11.07	383	69	2.30	5.92
Stroop congruity	77**		4.68**		24**		1.31**	

*$p < .05$. **$p < .01$.

$F(1, 18) = 65.76$, $p < .001$, $MSE = 142{,}528.34$. For number, the RT difference between congruent and incongruent pairs was 77 ms, $F(1, 16) = 260.87$, $p < .001$, $MSE = 1{,}127{,}632.5$, about three times the Stroop congruity obtained for size (24 ms), $F(1, 16) = 18.15$, $p < .001$, $MSE = 120{,}196.80$. The error data closely mirrored the RT data. Although incongruent pairs were more erroneous than congruent pairs for both number, $F(1, 16) = 33.06$, $p < .001$, $MSE = 4{,}143.71$, and size, $F(1, 16) = 20.95$, $p < .001$, $MSE = 342.08$, the difference was more than three times larger for number (4.68%) than for size (1.31%). The difference in magnitude of the Stroop effect for the two dimensions was highly significant, $F(1, 18) = 19.36$, $p < .001$, $MSE = 566.93$.

Discussion

In this experiment, we made size (a physical dimension) more discriminable than numerical magnitude (a semantic dimension). As a result, irrelevant size interfered with comparisons of number considerably, but irrelevant numerical magnitude affected comparisons of size to a lesser (yet still significant) degree. The failure of selective attention was much greater for the less discriminable dimension of number than for the more discriminable dimension of size.

The results of Experiment 5 are a mirror image of those found in Experiment 4. In Experiment 4, we made number more discriminable than size. As a result, the pattern of interference was asymmetrical, in favor of numerical value. In the current experiment, we altered the imbalance in favor of physical size and found a reversal of the pattern of interference in favor of size. These results underscore the crucial role of relative discriminability in shaping the dimensional interaction between number and size.

The results of Experiment 5 are compatible with views of partial automaticity, but they challenge theories positing strong automaticity for number. One should note that the numbers presented in this experiment were those used in the previous experiment. Nevertheless, the participants were better able to ignore their semantic component (their numerical magnitude) in Experiment 5 in which the number dimension was less discriminable than the dimension of size. We concluded that the activation of irrelevant numerical magnitude is strongly affected by attentional factors and is therefore not strongly automatic.

Experiment 6

To test whether the processing of number is even partially automatic, we planned two auxiliary experiments. Because the theoretical burden in this realm resides in the notion of automatic activation of semantic information, the participants in these experiments only performed comparisons of physical size. In Experiments 6 and 7, we thus tested whether comparison of size (a nonsemantic dimension of the numerals) was affected by irrelevant numerical magnitude (the semantic dimension of the numerals). If so, were the effects of irrelevant magnitude sensitive to practice and motivation? In Experiment 6, we tested whether the processing of irrelevant numerical information was affected by practice. The interference from irrelevant numerical value found in Experiment 3 might be dischargeable through extended practice. Consequently, in Experiment 6 the participants had four times as many practice trials as in the previous experiments. In Experiment 7 we examined whether activation of semantic information was additionally influenced by motivation. To augment the motivation of participants to attend to relevant size, we promised them a cash bonus proportional to the speed and accuracy of their performance. Practice was further extended: In Experiment 7, the participants went through three experimental sessions separated by at least one day. Sensitivity of the semantic intrusions to practice and motivation would further attest to the presence of considerable attentional and strategic influences.

Method

Participants. The participants were 8 Bar-Ilan undergraduates who had not participated in any of the previous experiments. Their ages ranged from 21 to 25 years. They were paid for their participation in the experiment.

Apparatus and stimuli. The stimulus pairs used were those from Experiments 2 and 3. We had already established in these experiments that the values on each dimension were equally discriminable at baseline. Apparatus, stimulus presentation, and viewing conditions were the same as those used in the previous experiments.

Procedure. The participants were instructed to compare the pairs of numerals on physical size by selecting the larger member of each pair as quickly and accurately as possible. Each person participated in a single experimental session. Half the participants performed the baseline task first, and half performed the filtering task first. In each task, the participants first went through 160 practice trials, four times the number in Experiment 3. Following this extended practice, they then proceeded to complete the experimental trials. They could rest after completing each block of 40 trials. An experimental session consisted of 320 practice trials and 320 experimental trials and lasted about 40 min.

Results

The correlation between speed and accuracy was $-.65$ ($p > .05$), indicating no significant trade-off between these two measures of performance. The summary of performance is presented in Table 5. Participants performed faster at baseline than in filtering, resulting in a significant Garner interference of 21 ms, $F(1, 7) = 5.89$, $p < .05$, $MSE = 13{,}302.13$. For errors, however, the difference between the two tasks (0.69%) was not significant, $F(1, 7) = 1.45$, $p > 0.2$. For Stroop effects, performance on congruent trials was both faster and more accurate than on incongruent trials. Stroop congruity was 53 ms for speed, $F(1, 7) = 45.62$, $p < .001$, $MSE = 91{,}800.23$, and 3.81% for accuracy, $F(1, 7) = 22.45$, $p < .005$, $MSE = 463.03$.

Discussion

In Experiment 6, Garner interference was fairly small for both RT and error; indeed, it was insignificant for the latter. However, substantial Stroop congruity was found for both speed and accuracy. Consider the current results in conjunction with those obtained for comparisons of size in Experiment 3. Although baseline RTs were shorter in Experiment 3 than in Experiment 6 (431 and 448 ms, respectively), both Garner interference and Stroop congruity were appreciably smaller in Experiment 6 than in Experiment 3 (by 19 ms and 12 ms, respectively). For error, both Garner (0.69%) and Stroop (3.81%) effects in the current experiment were approximately half the respective values in Experiment 3 (1.97% and 7.33%). For errors, Garner interference was significant in Experiment 3 but was not in the current experiment.

We concluded that unintentional activation of semantic numerical information is affected considerably by practice. Compared with Experiment 3, in which minimal practice was provided, the extended practice in Experiment 6 reduced interference from magnitude representation appreciably. Nevertheless, by no means was the activation of a magnitude representation eliminated in the current experiment. Therefore, the hypothesis of partial automaticity in the processing of numerical information is supported by the results of this experiment.

Table 5
Mean Reaction Times (RTs; in Milliseconds) and Error Rates for Comparison of Physical Size Across Task (Garner Interference) and Pair Type (Stroop Congruity) in Experiment 6

Assessment of Garner and Stroop effects	RT		Error (%)	
	M	*SD*	*M*	*SD*
Garner analyses				
Task				
Baseline	448	46	2.10	4.58
Filtering	469	59	2.79	6.16
Garner interference	21*		0.69	
Stroop analyses				
Pairs				
Congruent	442	46	0.88	2.66
Incongruent	495	59	4.69	7.88
Stroop congruity	53**		3.81**	

*$p < .05$. **$p < .01$.

Experiment 7

Method

Participants. The participants were 8 Bar-Ilan University undergraduates who had not taken part in any of the previous experiments. Their ages ranged from 20 to 24 years. They were paid for their participation in the experiment.

Apparatus and stimuli. The stimulus pairs were those used in Experiment 6. Apparatus, stimulus presentation, and viewing conditions were the same as those of the previous experiments.

Procedure. Again, the participants were instructed to compare the physical size of the pairs of numerals. Each person participated in three experimental sessions, separated by at least 24 hr. At the onset of the first session, participants were promised a cash bonus proportional to the speed and accuracy with which they performed the experimental task. With the exception of these instructions, the experimental sessions followed the plan of the previous experiment. Again, the first half of the trials were devoted to extended practice. Following the completion of all three sessions, all participants were told that their performance was excellent and were paid the maximal bonus.

Results

The correlation between speed and accuracy was $-.76$ ($p < .05$), indicating a trade-off between the two measures under the current conditions. Again, performance was faster at baseline than in filtering, which indicates a significant Garner interference of 19 ms, $F(1, 7) = 49.93$, $p < .001$, $MSE = 33{,}396.93$. For error rate, however, the difference between the two tasks (0.46%) did not amount to a

significant Garner interference, $F(1, 7) = 1.23, p > .3$. The data for congruent and incongruent pairs are also presented in Table 6. Performance for congruent pairs was superior by 36 ms in speed, $F(1, 7) = 41.09$, $p < .001$, $MSE = 123{,}106.90$, and by 2.96% in accuracy, $F(1, 7) = 9.84, p < 0.05, MSE = 840.40$.

Examining the Stroop effect separately for each session, we found that it declined appreciably across sessions. As shown in Figure 2, Stroop congruity decreased from 43 ms in the first session to 26 ms in the last. The decrease in magnitude was significant, $F(2, 14) = 6.13, p < .05, MSE = 2{,}433.06$, although the Stroop effect was dependable even in the third session, $t(7) = 4.92, p < .005$. However, given the steep slope of decline of the Stroop function, it might be interesting to speculate that additional practice would have reduced the effect further or possibly eliminated it.

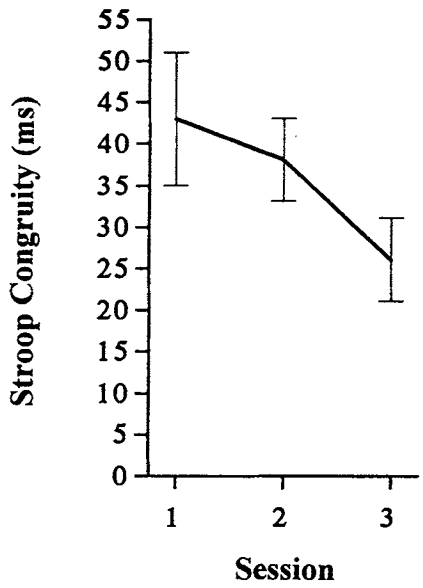

Figure 2. Stroop congruity as a function of consecutive experimental sessions (Experiment 7). The vertical bars mark one standard error around the mean.

Discussion

In Experiment 7, irrelevant variation in number somewhat impaired the speed with which the participants compared size, but it did not affect their accuracy. Stroop congruity was significant for both RT and error. Notably, the Stroop effect decreased appreciably with practice; the shrinkage derived from reduced intrusions from irrelevant magnitude. With practice, the participants were better able to withstand numerical interference and to attend selectively to physical size. Our results again show that, at best, the activation of numerical value is only partially automatic.

When the current judgments of size were compared with those of Experiments 3 and 6, the small interference found in Experiment 6 was further reduced in Experiment 7. In particular, Stroop congruity declined from 65 ms in Experiment 3 to 53 ms in Experiment 6 to 36 ms in the current experiment. The results of this experiment show that the interference caused by irrelevant numerical magnitude can be reduced, if not wholly eliminated, by factors such as practice and motivation.[3]

Table 6
Mean Reaction Times (RTs; in Milliseconds) and Error Rates for Comparison of Physical Size Across Task (Garner Interference) and Pair Type (Stroop Congruity) in Experiment 7

Assessment of Garner and Stroop effects	RT		Error (%)	
	M	*SD*	*M*	*SD*
Garner analyses				
Task				
Baseline	416	53	1.60	3.29
Filtering	435	62	2.06	5.08
Garner interference	19**		0.46	
Stroop analyses				
Pairs				
Congruent	417	54	0.58	2.05
Incongruent	453	65	3.54	6.57
Stroop congruity	36**		2.96*	

*$p < .05$. **$p < .01$.

General Discussion

The Role of Attention in the Perception of Number

The seven experiments of this study established the role of attention in regulating the extent of semantic involvement in numerical perception. The degree to which irrelevant semantic (magnitude) information intrudes on processing of physical size (a nonsemantic attribute of numerals) was shown to be malleable experimentally in a straightforward fashion. Both the Garner and Stroop effects obtained for judgments of size varied systematically across conditions, ranging from minuscule (good selective attention) to sizable (failure of selective attention) amounts. The same pattern obtained for judgments of number as well, but theoretical interest has mainly focused on the former: The extent to which judgments of size are afflicted by irrelevant magnitude betrays the scale and nature of semantic activation in numerical perception. The results of the current study show the magnitude of that activation to depend lawfully on a few stimulus factors affecting attention.

[3] We encouraged motivation for optimal performance in Experiment 7 by promising the participants a cash bonus proportional to the quality of their performance. The influence of this manipulation alone on the ability to ignore irrelevant magnitude may be assessed by comparing performance in the first experimental session with that in Experiment 6. Stroop congruity was smaller in the first session of Experiment 7 (43 ms) than in Experiment 6 (53 ms); the difference indicated the effect of motivation alone on selective attention.

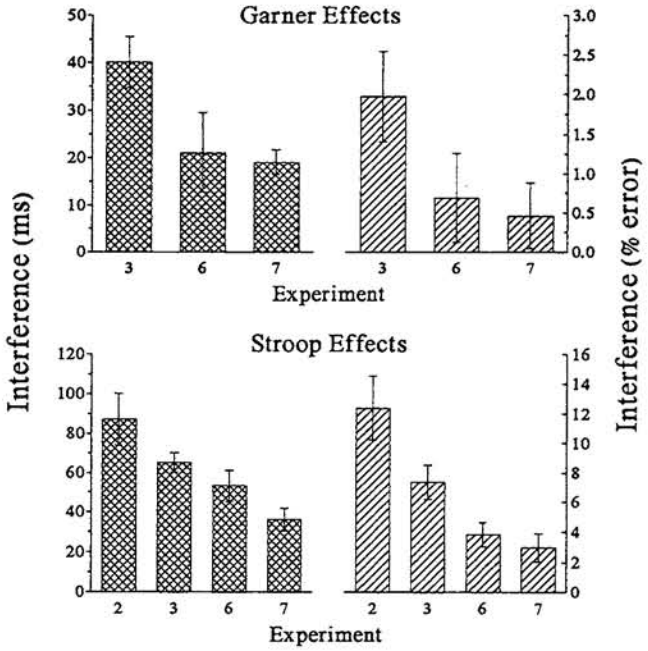

Figure 3. Garner (top) and Stroop (bottom) effects for judgments of physical size across experiments, for RT (left panel) and error rate (right panel). The vertical bars mark one standard error around the mean. In Experiment 2, all types of stimulus pairs were presented in a single block of trials. Practice in judging size was quadrupled from Experiments 2–3 to Experiment 6. The number of experimental sessions was tripled from Experiment 6 to 7. In Experiment 7, incentive motivation was introduced to induce attention to relevant size. Interference from irrelevant variation in numerical magnitude (Garner interference) as well as interference from conflicting numerical magnitude (Stroop interference) decreased in a monotonic fashion across experiments.

The dependence of semantic intrusion on practice and motivation is depicted in Figure 3. Judgments of size suffered more interference from irrelevant magnitude in Experiments 2 and 3 than in Experiments 6 and 7. Quadrupling practice in Experiment 6 reduced Garner and Stroop interference appreciably for both speed and accuracy. Extending practice further and augmenting motivation in Experiment 7 depressed Stroop and Garner effects to less than half their original values. The sensitivity of both measures to changes in practice and motivation attests to the presence of considerable strategic involvement in numerical perception.

Perhaps the strongest piece of evidence for the malleability of a magnitude representation for number comes from the examination of the collective results of Experiments 3–5. As shown in Figure 4, both Garner and Stroop effects changed in a lawful manner as a function of dimensional discriminability. When discriminability favored number (Experiment 4), comparisons of number were less afflicted by intrusions from irrelevant size than were comparisons of size from intrusions of irrelevant number. When the dimensional imbalance was tilted in favor of size (Experiment 5), the pattern of interference was reversed. A notable feature of the collective data depicted in Figure 4 is the complete symmetry of the criterial dimensions in response to the manipulation of discriminability. The dimension of number was as sensitive to relative discriminability as was the dimension of size; the former did not enjoy a processing privilege as a result of its semantic nature.

These results demonstrate the extent to which semantic and physical processing of numerals can be brought under

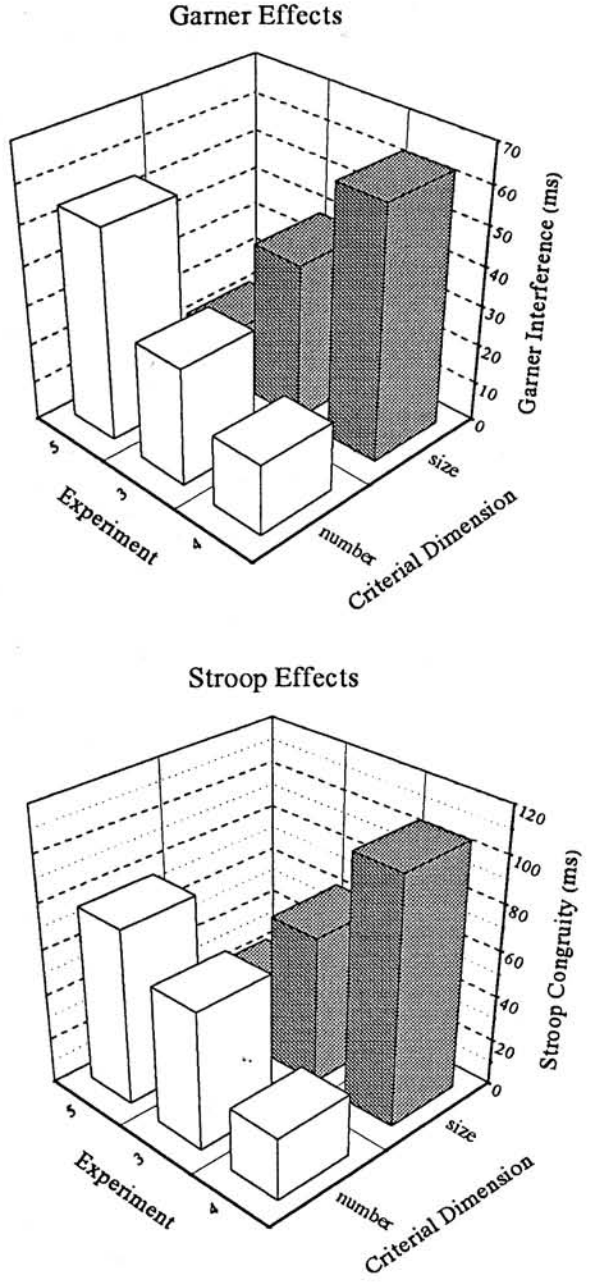

Figure 4. Garner (top) and Stroop (bottom) effects for comparisons of number and size in three experiments in which there was different relative discriminability of the two dimensions. In Experiment 3, the dimensions were equally discriminable; in Experiment 4, number was more discriminable than size; and in Experiment 5, size was more discriminable than number.

stimulus control. The quality of selective attention to physical size (or numerical magnitude) depends on a few experimental factors, the most important of which are (a) relative discriminability and (b) information value (set size) of the stimuli along the two dimensions. By judiciously manipulating these factors, we made one or the other dimension of numerals more salient or informative. That dimension, in turn, captured attention and was better able to withstand intrusions from an irrelevant dimension that was also less salient. Variations in the effectiveness of dimensional processing thus depended lawfully on the allocation of attention.

The potency of the attentional control revealed in the current results argues against claims of strong automaticity in numerical perception. Fully automatic processes are unaffected by attention, an outcome that clearly did not hold in the current case. Our results are more compatible with views of partial automaticity. In that view, semantic processing of numerals is ubiquitous, but it is nonetheless modulated by attention. Commensurate with partial or weak automaticity, we note the resilience of Stroop (and Garner) effects throughout the various conditions of this study. Although Stroop congruity varied from small to large across experiments, a statistically significant effect appeared in virtually every case. Even in Experiment 7, in which the dimensions were matched in discriminability and participants practiced for 3 days and were given a cash bonus, judgments of physical size produced a significant effect of Stroop congruity. The current results, which support the view of partial automaticity, demonstrate the effects of attention over and above those produced by automatic processing.

The Case Against Automaticity

Nevertheless, can an argument be built against the notion of automatic activation of numerical magnitude, full *or* partial? We believe that several features of the current data conspire to challenge even the notion of partial automaticity for number. Foremost among these is the absence of a Stroop effect for judgments of size in Experiment 1. This is a striking result that shows that under certain circumstances, people can process (physical attributes of) numerals without suffering interference from their semantic components.[4]

Additional evidence for the conclusion that numerical Stroop effects are not universally obligatory comes from detailed inspection of portions of the comparison data. Apart from the omnibus Stroop effect calculated for an experiment (say, for judgments of size), one can calculate a Stroop effect for each pair of sizes (or numbers) included in the experiment. The results of such a microanalysis, plotting Stroop congruity against the distance separating the two stimuli in a pair, are shown in Figure 5 for the size judgments of Experiment 5. Recall that size was more discriminable than number in that experiment, yet it, too, produced a (global) Stroop effect. However, the microanalysis shows Stroop congruity to decrease with distance, $F(4, 64) = 3.64$, $p < .01$, $MSE = 4{,}178.31$, to 10 ms at the largest value of intrapair separation. For the latter stimuli—pairs in which the asymmetry favoring size was greatest—Stroop congruity vanished, $t(19) = 1.26$, $p < .2$. In other words, for those eminently discriminable stimuli, comparisons of size were free of intrusions from irrelevant magnitude.

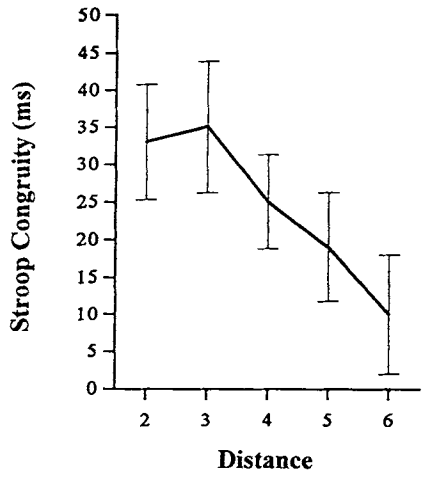

Figure 5. Stroop congruity for comparisons of physical size for each value of distance separating the two stimuli in a pair (Experiment 5). The vertical bars mark one standard error around the mean.

Let us put these observations in perspective. We already noted the great resilience of Stroop and Garner effects in the face of tight stimulus controls. What is really intriguing in our data is how very difficult it was to get rid of them. Consequently, we concluded that the overall pattern of the results best corresponded to the idea of partial automaticity. That granted, one must acknowledge several trends in the

[4] Additional pieces of evidence speak to the contingent nature of semantic processing in numerical perception. An important one is the absence of spatial–numerical associations in large portions of the physical comparison data. Dehaene et al. (1993) and Dehaene and Akhavein (1995) have reported strong spatial–numerical associations for odd–even, same–different, and parity judgments of numerals (i.e., judgments characterized by being unrelated to magnitude); large numbers were responded to more quickly when they appeared at the right-hand side of the extracorporal space, and small numbers were responded to more quickly when they appeared on the left-hand side. Dehaene and his associates have interpreted the spatial–numerical association of response codes (SNARC) effects found as evidence for the mandatory activation of semantic processing even when magnitude information is irrelevant to the task at hand. By the same token, our failure to detect the SNARC effect should be taken as evidence against the automatic activation of semantic processing. We examined number–space associations in each experiment, separately for each criterial dimension. Except for a 16-ms trend in Experiment 1 (not very consequential given the fact that only two numbers were used), in no case was the SNARC effect significant. The absence of the SNARC effect is especially notable for comparisons of physical size.

data that cast doubt on accounts of mandatory activation of magnitude.

An ANOVA Model of Attention to Irrelevant Information

Apart from the quality of automatic processing, the current results underscore the role of attention in the ability of people to ignore irrelevant information. In our model, we partitioned the various sources of task-irrelevant information. These sources (may) capture attention, taking a toll on performance with the relevant dimension. Traditionally, researchers have concentrated their interest on one harmful source of irrelevant information: conflicting values of the stimuli on the to-be-ignored dimension. Stroop is a within-stimulus interference effect, depending on attention to the *content* (corresponding or conflicting) of the to-be-ignored dimension (Pomerantz, 1983, 1991; see also Pomerantz, Pristach, & Carson, 1989). Note, however, that for stimuli to be congruent or incongruent (and for a Stroop effect to appear), the observer must notice the changes in the to-be-ignored dimension. Other sources of variation determine if the observer attends to the irrelevant dimension.

These additional sources of variation reside in the entire stimulus ensemble, not within the individual stimulus. The first refers to the packaging of the stimuli: Congruent, incongruent, and unidimensional stimuli may appear in one block or in separate blocks. The former arrangement is bound to invoke more attention to the irrelevant dimension than the latter. The second species of Garner interference refers to the size of the stimulus set. The experimental stimuli may be drawn from a large set of values, thereby exhibiting great trial-to-trial variation, or they may be drawn from a small set of values with little trial-to-trial variation. A stimulus drawn from a large set reduces more uncertainty, or conveys more information, than does a stimulus drawn from a small set. As a result, a many-valued irrelevant dimension captures more attention than does a little-valued one.

In our scheme, Stroop interference can only appear if there is Garner interference, although the reverse does not hold (see Pomerantz et al., 1989, who derived the same conclusion). Garner interference registers the attention paid to a task-irrelevant dimension resulting in the complementary failure of selective attention to the relevant dimension. Noticing the irrelevant dimension, congruent and incongruent stimuli may be responded to differently, producing the Stroop effect. Conversely, the absence of Garner interference means that variation along the irrelevant dimension was not noticed by the observer. As a result, congruent and incongruent stimuli have no psychological reality; therefore no Stroop effect is possible.

The hallmark of Garner's (1974) approach is the contention that performance with a given stimulus depends critically on the identity of the other stimuli that could have appeared on that trial. Take the neutral pair 6–4 (physical size held constant) as an example. The comparison of magnitude for that pair is performed better if the other pairs in that block of trials are neutral as well rather than if the other pairs form a mixture of neutral and (in)congruent pairs. The same comparison is also performed better if the digits between 3 and 7 can only appear in the block of trials than if the set of permissible values spans the range between 1 and 9. The current results support the validity of Garner's analysis.

Therefore, the key concept in our model is that of variation-produced attention. The greater the variation of values along an irrelevant dimension, the more difficult it becomes to ignore it. The reason is that increased variation makes each experienced stimulus along that dimension that much more informative. As a result, the dimension captures attention, leading to the mandatory failure of selective attention to the relevant dimension. Discriminability may also be seen as a means for increasing (or decreasing) variation. Incidentally, discriminability is orthogonal to the repertoire of the three sources of interference because they affect performance similarly for given levels of discriminability. Our manipulations thus affected the salience of the tested dimensions by altering the allocation of attention to those dimensions. By decomposing the sources of irrelevant variation, our model was able to predict the relative salience of the constituent dimensions and, consequently, the direction and magnitude of the respective amounts of interference.

References

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, Spain: International Society for Psychophysics.

Algom, D., Dekel, A., & Pansky, A. (1996). The perception of number from the separability of the stimulus: The Stroop effect revisited. *Memory & Cognition, 24,* 557–572.

Algom, D., & Pansky, A. (1993). Perceptual and memory-based comparisons of area. In A. Garriga-Trillo, P. R. Minon, C. Garcia-Gallego, C. Lubin, J. M. Merino, & A. Villarino (Eds.), *Fechner Day '93* (pp. 7–12). Palma de Mallorca, Spain: International Society for Psychophysics.

Ashby, F. G., & Maddox, W. T. (1994). A response time theory of perceptual separability and perceptual integrality in speeded classification. *Journal of Mathematical Psychology, 33,* 423–466.

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 and Performance, 2,* 435–447.

Besner, D., & Coltheart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. *Neuropsychologia, 17,* 467–472.

Buckley, P. B., & Gillman, C. B. (1974). Comparison of digits and dot patterns. *Journal of Experimental Psychology, 103,* 1131–1136.

Curtis, D. W., Paulos, M. A., & Rule, J. S. (1973). Relation between disjunctive reaction time and stimulus difference. *Journal of Experimental Psychology, 99,* 167–173.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44,* 1–42.

Dehaene, S., & Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 314–326.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and numerical magnitude. *Journal of Experimental Psychology: General, 122,* 371–396.

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 and Performance, 16,* 626–641.

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

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, and 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., Podgorny, P., & Frasca, E. M. (1982). Physical and cognitive dimensions in stimulus comparison. *Perception & Psychophysics, 31,* 507–522.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356–388.

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.

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

Hurford, J. R. (1987). *Language and number.* Oxford, England: Basil Blackwell.

Johnson, D. M. (1939). Confidence and speed in two-category judgment. *Archives of Psychology, 241,* 1–52.

Kahneman, D., & Chajczyk, D. (1983). Tests of the automaticity of reading: Dilution of Stroop effects by color-irrelevant stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 497–509.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 29–61). San Diego, CA: Academic Press.

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

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

Maddox, W. T., & Ashby, F. G. (1996). Perceptual separability, decisional separability, and the identification–speeded classification relationship. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 795–817.

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., & 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.

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.

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.

Posner, M. I. (1978). *Chronometric explorations of the mind.* 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 and Performance, 6,* 140–150.

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

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

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.

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 and Performance, 13,* 89–103.

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

Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory, and 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 and Language, 36,* 117–126.

Washburn, D. A. (1994). Stroop-like effects for monkeys and humans: Processing speed or strength of association? *Psychological Science, 5,* 375–379.

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

Received January 15, 1997
Revision received September 22, 1997
Accepted November 3, 1997 ■

Instructions to Authors

Journal of Experimental Psychology: Human Perception and Performance

Authors should prepare manuscripts according to the *Publication Manual of the American Psychological Association* (4th ed.). All manuscripts must include an abstract containing a maximum of 960 characters and spaces (which is approximately 120 words) typed on a separate sheet of paper. All copy must be double-spaced. Instructions for preparation of abstracts, tables, figures, references, and abbreviations of measurements appear in the *Manual.* All manuscripts are subject to editing for bias-free language (see chap. 2 of the *Publication Manual*). For further information on content, authors may refer to an editorial by a previous editor, James Cutting, in the May 1988 issue of the *Journal* (Vol. 14, No. 2, p. 322). For information on the other *JEP* journals, authors should refer to editorials in those journals.

APA policy prohibits an author from submitting the same manuscript for concurrent consideration by two or more publications. In addition, it is a violation of APA Ethical Principles to publish "as original data, data that have been previously published" (Standard 6.24). As this journal is a primary journal that publishes original material only, APA policy prohibits as well publication of any manuscript that has already been published in whole or substantial part elsewhere. Authors have an obligation to consult journal editors concerning prior publication of any data upon which their article depends. In addition, APA Ethical Principles specify that "after research results are published, psychologists do not withhold the data on which their conclusions are based from other competent professionals who seek to verify the substantive claims through reanalysis and who intend to use such data only for that purpose, provided that the confidentiality of the participants can be protected unless legal rights concerning proprietary data preclude their release" (Standard 6.25). APA expects authors submitting to this journal to adhere to these standards. Specifically, authors of manuscripts submitted to APA journals are expected to have available their data throughout the editorial review process and for at least 5 years after the date of publication.

Authors will be required to state in writing that they have complied with APA ethical standards in the treatment of their sample, human or animal, or to describe the details of treatment. A copy of the APA Ethical Principles may be obtained by writing the APA Ethics Office, 750 First Street, NE, Washington, DC 20002-4242.

Masked reviews are optional, and authors who wish masked reviews must specifically request them when submitting their manuscripts. Each copy of a manuscript to be mask reviewed should include a separate title page with authors' names and affiliations, and these should not appear anywhere else on the manuscript. Footnotes that identify the authors should be typed on a separate page. Authors should make every effort to see that the manuscript itself contains no clues to their identities.

Five copies of the manuscripts should be submitted. All copies should be clear, readable, and on paper of good quality. A dot matrix or unusual typeface is acceptable only if it is clear and legible. In addition to addresses and phone numbers, authors should supply electronic mail addresses and fax numbers, if available, for potential use by the editorial office and later by the production office. Authors should keep a copy of the manuscript to guard against loss. Effective in January 1999, the Incoming Editor is receiving all submissions to the journal. Submissions that are accepted will be published beginning in the 2000 volume. Mail manuscripts to the Incoming Editor, David A. Rosenbaum, Department of Psychology, Pennsylvania State University, 642 Moore Building, University Park, Pennsylvania 16802-3104.

For the other *JEP* journals, authors should submit manuscripts (in quadruplicate) to one of the editors at the following addresses: *Journal of Experimental Psychology: General,* Nora S. Newcombe, Department of Psychology, Temple University, 565 Weiss Hall, Philadelphia, PA 19122; *Journal of Experimental Psychology: Learning, Memory, and Cognition,* James H. Neely, Department of Psychology, State University of New York at Albany, 1400 Washington Avenue, Albany, NY 12222; *Journal of Experimental Psychology: Animal Behavior Processes,* Mark E. Bouton, Editor, Department of Psychology, University of Vermont, Burlington, VT 05405; and *Journal of Experimental Psychology: Applied,* Raymond S. Nickerson, Department of Psychology, Tufts University, Medford, MA 02155. When one of the editors believes a manuscript is clearly more appropriate for an alternative journal of the American Psychological Association, the editor may redirect the manuscript with the approval of the author.