Journal of Experimental Psychology:
Learning, Memory, and Cognition
2002, Vol. 28, No. 2, 259–274

0278-7393/02/$5.00 DOI: 10.1037//0278-7393.28.2.259

Comparative Judgment of Numerosity and Numerical Magnitude: Attention Preempts Automaticity

Ainat Pansky
Bar-Ilan University

Daniel Algom
Tel Aviv University

It is commonly believed that humans are unable to ignore the meanings of numerical symbols, even when these meanings are irrelevant to the task at hand. In 5 experiments, the authors tested the notion of automatic activation of numerical magnitude by asking participants to compare, while timed, pairs of numerical arrays on either numerosity or numerical value. Garner and Stroop effects were used to gauge the degree of interactive processing. The results showed that both effects were sensitive to the discriminability of values along the constituent dimensions, to the number of stimulus values used, and to practice and motivation. Notably, Stroop and Garner effects were eliminated under several conditions. These findings are incompatible with claims of obligatory activation of meaning in numerical processing, and they cast doubt on theories positing automatic processing of semantic information for alphanumerical symbols.

The arrays shown in Figure 1A illustrate the stimuli used in this study. The task for the participant was to decide, while timed, which array contained more numerals (to keep area constant, asterisks were used as fillers). By the principle of cardinality (Frege, 1884/1980; Russell, 1903, 1919; see also, Brainerd, 1979), each element in the left-hand array of 8s denotes a larger referent collection than each element in the right-hand array of 2s. It is by this principle that people consider “8” to be larger than “2.” However, applying cardinality to the collections of numerals in Figure 1A, the left-hand array of 8s is smaller than the right-hand array of 2s. Therefore, the array of 8s is both larger and smaller than the array of 2s.

Does the paradox inhering in the stimuli of Figure 1A entail behavioral expressions, such that judgments of numerosity are affected by irrelevant numerical magnitude? If so, can the interference be avoided, or is it inescapable? Do humans (and animals; cf. Washburn, 1994) engage the meanings of numerical symbols in a mandatory fashion, even when the meanings are irrelevant to the task at hand and can hurt performance? The purpose of this study was to answer these fundamental questions. We used selective attention as a tool to uncover the nature of numerical processing.

Effects of Stroop (1935; see also MacLeod, 1991, 1992) and Garner (1974)[1] served as our means to gauge selectivity. For the latter effect, we compared numerosity performance in a condition in which there was trial-to-trial variation in irrelevant numerical magnitude (the Garnerian filtering task) with that in a condition in which irrelevant numerical magnitude was held constant (the Garnerian baseline task). Performance detriment in the filtering task—Garner interference—attests to the absence of full selective attention to numerosity. For the former effect, we measured the difference in performance between stimuli matched (the more numerous array containing the numerically larger numbers) and mismatched (as in Figure 1A) on the dimensions of numerical magnitude and numerosity. Impaired performance with the conflicting stimuli—Stroop congruity—further attests to the failure of selective attention[2] to numerosity. We measured Stroop and Gar-

Ainat Pansky, Department of Psychology, Bar-Ilan University, Ramat Gan, Israel; Daniel Algom, Department of Psychology, Tel Aviv University, Ramat Aviv, Tel Aviv, Israel.

Ainat Pansky is now at the Institute of Information Processing and Decision Making, University of Haifa, Mount Carmel, Haifa, Israel.

The current study is based on a doctoral dissertation carried out by Ainat Pansky and supervised by Daniel Algom at the Department of Psychology, Bar-Ilan University. This research was supported by Israeli Trustees Foundation Grant DC-7/17 awarded to Ainat Pansky and by Israel National Science Foundation Grant 93-58 awarded to Daniel Algom. We thank James Pomerantz, Tom Wallsten, and two anonymous reviewers for their insightful comments. We also appreciate the comments by Gordon Logan and Tim Rickard 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, Tel Aviv 69978, Israel. E-mail: algomd@freud.tau.ac.il

[1] Originally, Stroop and Garner effects were derived in tasks of speeded classification of single stimuli. However, the effects have been subsequently derived in tasks of comparing pairs of stimuli (Garner, 1988). In the numerical domain, in particular, Stroop effects have commonly been investigated in comparison (e.g., Henik & Tzelgov, 1982; Washburn, 1994) rather than in classification. One should recognize though that the genesis and behavior of the effects may differ in tasks of comparison (e.g., in comparison there is irrelevant variation both within a pair and between pairs). One reviewer, James Pomerantz, suggested a new term, “Stroop Type B,” to cover these interference effects. To avoid proliferation of terms, we use the classic nomenclature, but the reader should recognize that in this study we use variants of Stroop and Garner effects adapted for the task of comparative judgment.

[2] It is important to distinguish between attention and performance. By no means is the failure of selective attention universally associated with poorer performance. The common condition (e.g., Algom et al., 1996; Logan & Zbrodoff, 1979, 1998) in which the dimensional values are correlated over trials (with either congruent or incongruent stimuli pre-

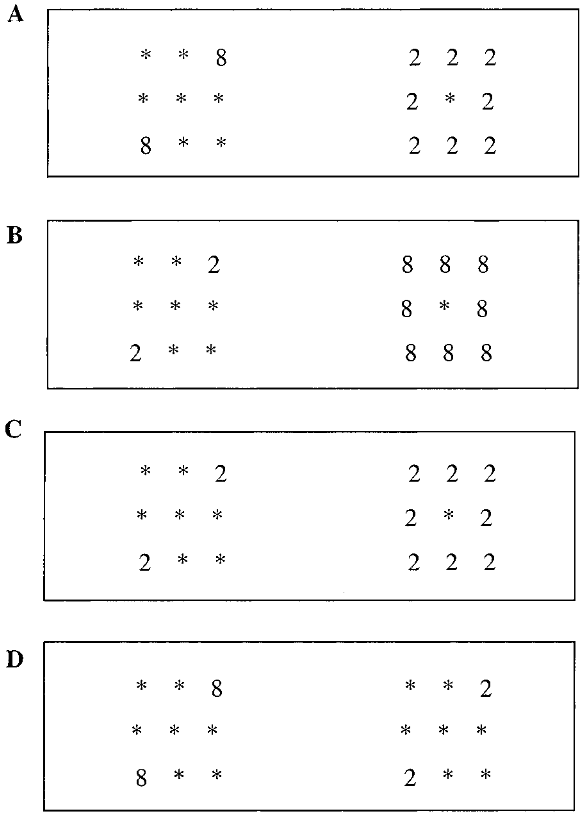

Figure 1. Examples of the stimulus displays used in this study. Panel A: an incongruent pair; Panel B: a congruent pair; Panel C: a neutral pair for comparisons of numerosity; and Panel D: a neutral pair for comparisons of numerical magnitude.

ner effects under distinct contexts that differed in the relative salience of the constituent dimensions. Given the documented power of context to affect selectivity, we expected the more salient dimension to interfere with performance on the less salient dimension, whether numerical magnitude or numerosity.

To foreshadow the results, interference—of both the Garner and the Stroop species—varied considerably across the experimental contexts. Under some contexts, performance was completely free of intrusions from the task-irrelevant dimension. Hence, interference is sometimes avoidable when people judge the numerosity of numerical symbols. We concluded that semantic (magnitude) information is not activated in a mandatory fashion just about whenever numerals are presented for any purpose. Under certain conditions, people can treat a numeral merely as an "ink mark on a piece of paper" (Stevens, 1951, p. 22).

The Activation of Numerical Magnitude

Our conclusions challenge a virtual unanimity favoring an automaticity account for numerical perception:

> the presentation of an arabic numeral elicits an automatic activation of the appropriate . . . magnitude code . . . [that] cannot be repressed, even though magnitude information is irrelevant to the task (Dehaene, 1992, p. 21)

> the magnitude of the digit may evoke a response through a fast, automatic process over which the subjects have little control (Sudevan & Taylor, 1987, p. 94)

> numerical distances are automatically computed (Henik & Tzelgov, 1982, p. 394)

> whenever we see a digit, its quantitative representation is immediately retrieved (Dehaene, 1997, p. 74)

> Stroop interference . . . in enumeration tasks depends on a rapid and automatic activation of digits' magnitude representation (Pavese & Umilta, 1999, p. 62).

Therefore, reading a number is considered to be automatic in the same sense that is usually attributed to the reading of words (cf. Besner, Stolz, & Boutilier, 1997; see Hurford, 1987, for an exposition of the shared linguistic nature of words and numbers):

> A fail-safe demonstration of automaticity, in particular the automatic nature of accessing word meaning, involves the Stroop task (Ashcraft, 1994, p. 72)

> the Stroop effect demonstrates that . . . the meaning of a word [is] processed by skilled readers even when they are trying hard not to process them (Rayner & Pollatsek, 1989, p. 72)

> Reading is such an automatic process that it is difficult to inhibit and it will interfere with processing other information about the word (Anderson, 1995, p. 100).

On the traditional view of the Stroop effect and reading, semantic information is called up ineluctably upon the presentation of alphanumerical symbols (see MacLeod, 1991, for a review). Reading a number is said to be automatic in the sense that people can not refrain from retrieving the meaning of the number (its magnitude) in the face of instructions to ignore it, even when semantic activation hurts performance (cf. Besner, 2001; Besner & Stolz, 1999; Kahneman & Treisman, 1984). It is this claim regarding the obligatory retrieval of magnitude information that is questioned in the current research.

Early accounts of automaticity (e.g., LaBerge & Samuels, 1974; Posner & Snyder, 1975; Shiffrin & Schneider, 1977) have construed it as processing without attention. Additional criteria included fast, ballistic, effortless, involuntary, unintentional, or autonomous responding (cf. Bargh, 1992; Hasher & Zacks, 1979; Posner, 1978; see also, Logan, 1988). Recent theories of automaticity (Bargh, 1992; Logan, 1988, 1991, 1992; Vallacher & Weg-

---

dominating in the set) provides a handy example. People notice the covariation, a feat only possible by attending to the values of the task-irrelevant dimension. By definition, exclusive attention to the relevant dimension has failed, and large amounts of the Stroop effect typically ensue under such contexts (see Dishon-Berkovits & Algom, 2000, and Shalev & Algom, 2000, for recent results). Simultaneously, performance (latency, errors) with the target dimension is very good, precisely because the irrelevant dimension was noticed. One should note the dissociation: Selective attention failed yet performance improved. When there is a large amount of facilitation, then (a) selective attention to the target dimension has failed, because, obviously, the nominally irrelevant dimension has been noticed, and (b) performance with the target dimension has improved precisely due to the failure of selectivity. Other relations between the two measures hold under various conditions; our point is that the two must be considered separately.

ner, 1987; see also, Boronat & Logan, 1997; Carr, 1992; Rickard, 1997) allow for other relations between attention and automaticity. In Logan's (1988) influential model—instance theory—automatic processing is postattentive (hence, dependent on attention), and Bargh's model (Bargh, 1992) makes automatic processing conditional on prior goals and attention. However, for all the changes in our understanding of automaticity, both early and recent models share the core assumption of mandatory activation of meaning. Our results, in contrast, do not accord well with the claim that semantic activation cannot be prevented.

We stress instead the role of attention in regulating the extent of semantic involvement in numerical perception. Attention, in turn, depends on the relative salience of the constituent dimensions. We predicted the less salient dimension to be particularly vulnerable to intrusions from the more salient dimension, regardless of whether numerical magnitude or numerosity was the more (or the less) salient dimension. By the same token, the activation of magnitude information is optional, subject to strategic influences associated with salience and concomitant attention. We manipulated salience by making the values on one or the other dimension more discriminable or uncertain, and hence, commanding priority in the allocation of attention.

The Role of Context in Numerical Cognition: The Effect of Dimensional Discriminability and Uncertainty

Our experiments build on the classic findings regarding contextual modulation of the selectivity of attention (Garner, 1962, 1974, 1983; Lockhead, 1966, 1972, 1979; Marks, 1987, 1989; Melara, 1992; Melara & Mounts, 1993; Pomerantz, 1983, 1986; Pomerantz, Pristach, & Carson, 1989; Posner, 1964; Shepard, 1964; see Lockhead & Pomerantz, 1991, for an overview). One factor of context—discriminability—specifies the size of the psychological differences separating stimulus values along a dimension (Algom, Dekel, & Pansky, 1993, 1996; Ben-Artzi & Marks, 1995, 1999; Garner, 1983; Garner & Felfoldy, 1970; Huettel & Lockhead, 1999; Melara & Mounts, 1993; Pomerantz, 1983; Sabri, Melara, & Algom, 2001; Shalev & Algom, 2000). To examine its role in numerical cognition, we (Algom et al., 1993, 1996) manipulated the discriminability of numerical and physical size of single digits, matching it in one condition, but rendering one or the other dimension more discriminable in other conditions. The more discriminable dimension disrupted selective attention to the less discriminable dimension (but not vice versa). Notably, when the dimensions were matched, Stroop and Garner effects vanished, and performance was free of interference for both numerical magnitude and physical size. We should be quick to add that one can find Stroop and Garner effects when discriminability on the two dimensions is matched, given the presence of other contextual biases.

Another factor of context—uncertainty—specifies the number of values per dimension. A many-valued stimulus dimension is more informative than a binary-valued dimension: The same stimulus is more surprising coming from the former than from the latter. A more informative dimension can better capture attention. For numerical and physical size of numerals, we (Pansky & Algom, 1999) found larger Stroop and Garner effects with a many-valued irrelevant dimension than with a binary-valued irrelevant dimension. Both discriminability and uncertainty were treated in the present series of experiments concerning the numerical magnitude and the numerosity of numerals.

For these dimensions of number, our reading of the pertinent literature indicates that, apart from a single notable exception (Flowers, Warner, & Polansky, 1979, treating the contextual factor of mode of responding), discriminability has been neither appreciated nor measured. Extant studies are characterized by a mismatch in discriminability favoring the dimension of numerical magnitude (e.g., Fox, Shor, & Steinman, 1971; Francolini & Egeth, 1980; Hock & Petrasek, 1973; Morton, 1969; Pavese & Umilta, 1998, 1999; Reisberg, Baron, & Kemler, 1980; Shor, 1971; Washburn, 1994; Windes, 1968; see also, Buckley & Gillman, 1974; Dehaene & Cohen, 1994; Garner, Podgorny, & Frasca, 1982). Given this form of dimensional imbalance, the trial-to-trial changes in numerical magnitude are experienced as being psychologically greater than the trial-to-trial changes in numerosity. The greater salience of numbers may undermine selective attention to numerosity. Indeed, the typical outcome in the literature is that of an asymmetry in interference favoring numerical magnitude: Judgments of nonarithmetic properties such as size or numerosity were disrupted by numerical magnitude, but judgments of magnitude were free of interference from the physical properties of the numerals. Theoretically, however, neither the imbalance nor its form is fixed or inevitable. Another relation may give rise to a different outcome, favoring numerosity in processing. This possibility was pursued in the present series of experiments.

The Present Study

Given the deep association of numerical magnitude (i.e., number) and numerosity in mathematics (Fraenkel, Bar-Hillel, & Levy, 1973; Frege, 1884/1980; Russell, 1903, 1919), philosophy (Hofstadter, 1979; Husserl, 1891), and psychology (Gallistel, 1989; Piaget, 1954, 1965), we elected to test these dimensions for (in)dependence in processing. Discriminability was matched (Experiments 1–2), mismatched in favor of numerical magnitude (Experiment 3), or mismatched in favor of numerosity (Experiment 4). In a final experiment (Experiment 5), we tested the extent of strategic influences on the Stroop effect through extended practice and incentive motivation. For uncertainty, we contrasted Experiment 1 with Experiments 2–5; the same dimensions were less informative in Experiment 1 than in Experiments 2–5. The observation of no Stroop effect or large variations thereof poses a challenge to the idea that the meaning of numerals is activated in a mandatory, ballistic fashion.

Experiment 1

The purpose of this experiment was to test the perceptual (in)dependence of the numerical magnitude and the numerosity of numerical symbols. The dimensions each assumed two values, and the stimuli were created from all possible combinations of these values. The participants made speeded comparisons of numerosity and of numerical magnitude (in separate blocks of trials). The design closely followed Garner's (1974) paradigm (originally developed for classification of single stimuli). Most important, baseline discriminability was matched for numerical magnitude and numerosity.

Method

Participants. The participants were 12 Bar-Ilan University undergraduates who were paid to perform in the experiment. Their ages ranged between 20 and 26 years.

Stimuli and apparatus. To avoid the confounding numerosity and space,[3] we placed the numerals in the cells of an imaginary 3 × 3 matrix, as illustrated in the patterns of Figure 1. Cells not occupied by numerals were filled with asterisks, keeping the area of the stimulus arrays constant throughout the experiment.

We used the numbers 2 and 8 for numerical magnitude, and the corresponding quantities of 2 and 8 for numerosity. A given array was created by reproducing a constant numeral (either 2 or 8) either 2 or 8 times. From the four unique arrays, all possible combinations of neutral, congruent, and incongruent pairs of stimuli were produced. In Figure 1A, we present an example of an incongruent pair. Notice that numerical magnitude is larger for the left-hand array in which numerosity is smaller. Presenting the 8s in the more numerous array and the 2s in the less numerous array creates a congruent pair (Figure 1B). Replacing the 8s with 2s (Figure 1C) or the 2s with 8s generates neutral pairs for comparison of numerosity, with numerical magnitude now held constant (at either 2 or 8). Presenting the 2s in one array and the 8s in the other, with numerosity held constant (at either 2 or 8), creates the neutral pairs for comparison of numerical magnitude (Figure 1D). For a given pair, the larger stimulus (on the criterial dimension) appeared at each location (right or left) on a random half of its presentations.

From the entire set of stimulus pairs, we created four experimental tasks (two entailing judgments of numerical magnitude, two entailing judgments of numerosity). For comparison of numerical magnitude, the participants performed in a baseline task (in two blocks of 20 neutral trials, with numerosity held constant at 2 in one block and at 8 in another block), and in a filtering task (40 trials with numerosity varying in an orthogonal fashion, the pairs being either congruent or incongruent). It should be noted that for both baseline and filtering, each stimulus was repeated 10 times. The participant performed in the same two tasks comparing numerosity. To recap, in the baseline task, participants compared the stimuli on the criterial dimension (numerical magnitude or numerosity) with the irrelevant dimension held constant. In the filtering task, the participants again compared the arrays on the criterial dimension, but the stimuli also varied on the irrelevant dimension.

On the basis of pilot testing, we determined the physical size of the numerals, asterisks, and arrays used, such that the speed and accuracy of the comparisons along the two criterial dimensions would match at baseline. Each numeral was 27 × 18 pixels large, a pixel extending 0.4 mm in any orientation. A numeral subtended 0.77° of visual angle in length and 0.52° in width. The size of an asterisk was 9 × 6 pixels, subtending a visual angle of 0.26° in length and 0.17° in width. An entire array (of nine stimuli) subtended approximately 3.00° of visual angle in both length and width. The stimuli were generated in Pascal-small font by an IBM-compatible (PC-386) microcomputer and displayed on a super-VGA (video graphics array) 14-in color monitor. The stimuli appeared white over a dark screen background at the centers of the left and right hemifields; to avoid adaptation, we introduced a trial-to-trial spatial uncertainty of up to five 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 performed the two numerical and two numerosity comparison tasks together as a set, with half of the participants first comparing numerical magnitude, and half first comparing numerosity. 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 sequence. Intervals of approximately 2 min separated the various tasks.

For comparisons of numerical magnitude, the participants were instructed to select the stimulus array containing the numerically larger elements. For comparisons of numerosity, they were instructed to select the stimulus array containing more numerals. The participants were encouraged to respond as quickly and as accurately as possible. Comparisons were made by pressing either a right- or left-hand key on the keyboard. The participant's task was to press the key on the side on which the larger member of the pair (according to the criterial dimension) appeared. The stimuli were response terminated. Each trial was presented following a 0.5-s pause after a response was given. Reaction time (RT) was measured in milliseconds using a software timer.

Data analysis. The experiments reported here were analyzed using the same general procedure. Trials in which the participant's RT was greater by more than three standard deviations than her or his mean for that type of trial (congruent, incongruent, or neutral) and judgment (numerical magnitude, numerosity) were excluded from the analyses. Planned comparisons among pairs of conditions were performed using the Bonferroni correction (.05 criterion value). Two global analyses of variance (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 numerosity, 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; order of dimensional judgment (numerical magnitude first, numerosity first) and task order (baseline first, filtering first) served as between-subject factors. For Stroop congruity in the filtering task, pair type (congruent, incongruent) and spatial organization (larger stimulus on the left or on the right) served as within-subject factors; order of dimensional judgment (comparisons of numerical magnitude first, comparisons of numerosity first) and task order (baseline first, filtering first) served as between-subject factors. In Experiments 2–5, ordinal distance served as an additional within-subject factor in each of the ANOVAs.

In each experiment, the correlation between RT and error rate was calculated for each participant across the various conditions. With the exception of Experiment 1 (Pearson *r* averaging .00), the coefficients of correlation were positive and fairly large (Pearson *r* averaging .63, .51, .61, and .49, respectively, in Experiments 2–5). There is no indication of a speed–accuracy trade-off in the current data. Instead, error and latency largely corresponded across the experimental conditions (see the respective pairs of means in Tables 1–5). Response times are typically more stable than accuracy (the latter, often, nearly perfect) in studies of speeded classification (Melara & Marks, 1990; Melara & Mounts, 1993) and speeded comparison (Petrusic & Baranski, 1989), wherefore we emphasize RT in our discussions.

[3] Numerosity is notorious for its interactions with other stimulus dimensions associated with numerical magnitude or quantity. Foremost among the latter is area, because the larger a collection of elements, the larger the space it usually occupies. Many studies (e.g., Bever, Mehler, & Epstein, 1968; Mehler & Bever, 1967; Piaget, 1965; Pufall & Shaw, 1972) have traced the development of the capacity to maintain perception of numerosity invariant in the face of changes in the irrelevant property of the space that the objects occupy. Full conservation of numerosity may not be attainable, though, because adults, too, have been shown to be affected by the spatial properties of a given collection of items (e.g., Dixon, 1978; Krueger, 1972; Vos, van Oeffelen, Tibosch, & Allik, 1988). Precisely to avoid such biases, we kept the total area of the numerical displays constant throughout all the experiments of the current study.

Table 1
Mean Reaction Times (RTs; in ms) and Percent Errors (PEs) for Comparison of Numerical Magnitude and Numerosity Across Task (Garner Interference) and Across Pair Type (Stroop Congruity) in Experiment 1

Task/stimuli	Numerical magnitude RT	Numerical magnitude PE	Numerosity RT	Numerosity PE
		Garner analyses		
Baseline task	386 (36)	2.76 (3.96)	390 (39)	3.79 (4.55)
Filtering task	395 (41)	0.86 (2.88)	405 (49)	1.87 (4.45)
Garner interference	9*	−1.90	15	−1.92
		Stroop analyses		
Congruent pairs	373 (33)	0.88 (2.98)	401 (51)	1.25 (3.38)
Incongruent pairs	416 (38)	0.83 (2.82)	408 (48)	2.50 (5.32)
Stroop congruity	43**	−0.05	7	1.25

Note. Standard deviations appear in parentheses. Error differences refer to percentage point effects (e.g., an increase in error rate from 1% to 2% is a 1 percentage point increase but a 100% increase in errors). The same treatment of error applies to the data presented in Tables 2–5.
* $p < .05$. ** $p < .01$.

Results

In Table 1 we present a summary of the data—speed of responding and proportion of errors—for comparisons of numerical magnitude and numerosity. For the former, average baseline performance was 386 ms; for the latter, it was 390 ms, a mere 4-ms difference, $t(11) = 0.60$, $p > .5$. The 1.03 percentage-point difference in baseline error rate, $t(11) = 1.42$, $p > .10$, provided additional evidence of matched discriminability.

To examine the selectivity of attention, we compared performance at baseline (in which the irrelevant dimension was held constant) with that in filtering (in which the irrelevant dimension varied in an orthogonal fashion). Garner interference, as shown in Table 1, was small for both judgments of numerical magnitude (9 ms), $F(1, 8) = 7.34$, $.01 < p < .05$, $MSE = 131.22$, and numerosity (15 ms), $F(1, 8) = 4.93$, $p > .05$, $MSE = 514.54$. For error, no Garner interference affected comparison performance for either numerical magnitude (−1.90%), $F(1, 8) = 3.96$, $p > .05$, $MSE = 11.01$, or numerosity (−1.92%), $F(1, 8) = 3.88$, $p > .05$, $MSE = 11.39$, as participants actually made fewer errors in filtering than at baseline.

In Table 1, we also present values of Stroop congruity for each dimension—the difference in performance between congruent and incongruent pairs of stimuli. For numerical magnitude, congruent pairs were compared 43 ms faster than incongruent pairs, $F(1, 8) = 143.59$, $p < .01$, $MSE = 156.58$. For numerosity, the RT advantage for congruent pairs was a minuscule 7 ms, $F(1, 8) = 2.21$, $p > .10$, $MSE = 231.28$. The dimensions thus differed with respect to the Stroop effects found, $F(1, 10) = 29.91$, $p < .01$, $MSE = 135.49$. For accuracy, no Stroop effect obtained for either numerical magnitude ($F < 1$) or numerosity, $F(1, 8) = 4.50$, $p > .05$, $MSE = 4.17$.

Discussion

The most important result to emerge from this experiment is the absence of Garner and Stroop interference in comparisons of numerosity. The absence of the former effect shows that our participants judged numerosity as speedily and accurately in a condition in which the numbers were held constant as they did in a condition in which the numbers varied across arrays and trials in a random fashion. The absence of a Stroop effect means that the participants judged pairs of stimuli in which the irrelevant numerical magnitude conflicted with numerosity as speedily and accurately as they judged pairs in which numerical magnitude matched numerosity. Collectively, the absence of Garner and Stroop effects shows that people can ignore numerical magnitude when judging numerosity.[4] Because they ignored irrelevant magnitude, the participants did not reap gain from corresponding numerical magnitude, nor did they suffer interference from conflicting numerical magnitude. This subset of the data is inconsistent with theories positing the mandatory activation of numerical magnitude whenever numerals are presented for view. Alternatively, one can maintain that numerical magnitude is always activated, but the activation was too weak to affect performance. Subsequent analyses of negative priming (see General Discussion) revealed that such was not the case.

Why have earlier studies reported Stroop interference for numerosity when we have found none? One explanation implicates the mode of responding: oral (mostly) in the former, manual in this study. Another explanation, one that we endorse, concerns relative dimensional discriminability. Baseline discriminability of numer-

[4] The one truly discordant result of this experiment is the large amount of reverse Stroop effect obtained for judgment of numerical magnitude (Garner interference was also significant for numerical magnitude, but its amount was too small to seriously compromise performance). The import of this result is modulated somewhat by the absence of both Garner and Stroop effects when we consider accuracy. Nevertheless, the reverse Stroop effect obtained for RT shows that our participants noticed numerosity when judging numerical magnitude, reaping gain from correspondence or incurring cost from conflict. We do not have a ready explanation for this discordant feature of the data.

ical magnitude and numerosity has been seldom measured or matched. As a result, the reported effects of Stroop might have been caused, in part, by an asymmetry in the baseline discriminability of the constituent dimensions. The more discriminable dimension (numerical magnitude) intruded on the less discriminable dimension (numerosity).

Experiment 1 differs in yet another detail from comparable studies in the literature. Previous experiments have typically presented multiple values of numerical magnitude, whereas we used a pair of values per dimension. Melara and Mounts (1994) have shown that the mere number of stimuli on an irrelevant dimension can affect classification performance on the relevant dimension. Kanne, Balota, Spieler, and Faust (1998) have shown that increases in stimulus set size are associated with larger Stroop effects (see Pansky and Algom, 1999, for a comparable association in numerical Stroop effects). Consequently, in the next experiment, we presented multiple pairs of stimuli, with many values of numerical magnitude and as many values of numerosity. Again, we matched the discriminability of the magnitudes and numerosities presented.

Experiment 2

We expanded the stimulus set to include many values of both numerical magnitude and numerosity. We included an equal number of values for both dimensions; we selected the values such that equal "psychological distances" separated a given pair of stimuli along the two constituent dimensions. For example, the speed and accuracy for comparing the numerals *3* and *7* (with numerosity held constant at either 3 or 7; numerical baseline task) were approximately the same as those of comparing the numerosities *3* and *7* (with numerical magnitude held constant at either 3 or 7; numerosity baseline task).

Method

Participants. Twenty Bar-Ilan University undergraduates were paid to participate. Their ages ranged from 21 to 25 years. None had taken part in the previous experiments.

Stimuli and apparatus. The values between 2 and 8 were used for both numerical magnitude and numerosity. All pairwise combinations of these values were selected; however, we excluded pairs containing adjacent values of numerical magnitude and numerosity, because we found it difficult to match some of these pairs in discriminability. The following 15 pairs were used: (2,4), (2,5), (2,6), (2,7), (2,8), (3,5), (3,6), (3,7), (3,8), (4,6), (4,7), (4,8), (5,7), (5,8), (6,8).

Take the pair of values (5,8) as an example. For judgments of numerical magnitude, the neutral pairs comprised arrays containing an equal number of numerals (either both arrays contained 5 numerals or both contained 8 numerals). For judgments of numerosity, the neutral pairs comprised arrays containing the same numeral (either both arrays contained only 5s or both contained only 8s). On either criterial dimension, one array contained 5 repetitions of the numeral 5, and the other contained 8 repetitions of the numeral 8, for congruent pairs. For incongruent pairs, one array contained 5 repetitions of the numeral 8, the other contained 8 repetitions of the numeral 5. Each pair was presented twice, with the larger stimulus appearing once on the right and once on the left. Similar combinations of congruent, incongruent, and neutral pairs were formed for all the other 14 pairs of values used.

Pilot testing confirmed that setting the size of each numeral at 27×18 pixels and that of each asterisk at 9×6 pixels, as in Experiment 1, would ensure equal baseline discriminability along the constituent dimensions. Thus, for each of the 15 pairs of values on each dimension, the speed and accuracy with which their two members were discriminated on numerical magnitude (with numerosity held constant) equaled the speed and accuracy with which their members were discriminated on numerosity (with numerical magnitude held constant).

The method was that used in Experiment 1. For each criterial dimension, the baseline task contained the pairs in which the value of the irrelevant dimension was held constant (neutral or unidimensional pairs). The filtering tasks contained the pairs in which the irrelevant dimension varied, with values either corresponding (congruent pairs) or conflicting (incongruent pairs) with those of the relevant dimension.

In Experiment 1 (as well as in mainstream Garner research), a single pair of values per dimension was used. In this experiment, both the baseline and the filtering tasks included multiple pairs of stimuli. Therefore, both the baseline and the filtering tasks entailed trial-to-trial variation in the values of the attended dimension, a feature not present in Experiment 1. The filtering task also entailed within-pair variation on the unattended dimension. Apparatus, stimulus presentation, and viewing conditions were those of the previous experiment.

Procedure. Each participant was tested in two (identical) experimental sessions, separated by at least 24 hr. In each session, half of the participants compared numerical magnitude first, and half compared numerosity first. For each criterial dimension, half of 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 60 trials as practice. They then proceeded to perform two blocks of the 60 trials, making a total of 120 judgments per task. The participants could rest for as long as needed between blocks. An entire experimental session, consisting of 240 practice trials and 480 experimental trials, lasted about 40 min.

Results

The results are shown in Table 2. Average baseline RT was 468 ms for comparisons of numerical magnitude and 457 ms for comparisons of numerosity, $t(19) = 1.77$, $p > .05$. Baseline error rates were also comparable for the two dimensions, differing by a mere percentage point of 0.32%, $t(19) = 0.85$, $p > .40$. The comparable performance at baseline for the two dimensions confirms our success at matching discriminability.

The magnitude of Garner interference was 19 ms for comparisons of numerical magnitude, $F(1, 16) = 10.49$, $p < .01$, $MSE = 6{,}140.67$, and 20 ms for comparisons of numerosity, $F(1, 16) = 10.25$, $p < .01$, $MSE = 6{,}738.40$. Error rates were also larger in filtering than at baseline. The difference was 2.21% for comparisons of numerical magnitude, $F(1, 16) = 11.16$, $p < .01$, $MSE = 56.64$, and 1.18% for comparisons of numerosity, $F(1, 16) = 4.27$, $p > .05$, $MSE = 61.20$. Random variation on the irrelevant dimension thus intruded on performance with the relevant dimension, whether numerical magnitude or numerosity. For Stroop effects, the speed advantage for congruent over incongruent pairs was 68 ms for numerical magnitude, $F(1, 16) = 102.01$, $p < .01$, $MSE = 6{,}631.47$, and 42 ms for numerosity, $F(1, 16) = 67.73$, $p < .01$, $MSE = 4{,}714.72$. Accuracy was also better for congruent than for incongruent pairs. The difference in error percentage point was 7.23% for comparisons of numerical magnitude, $F(1, 16) = 20.27$, $p < .01$, $MSE = 323.32$, and 3.88% for comparisons of numerosity, $F(1, 16)=11.50$, $p < .01$, $MSE = 249.43$. As in Experiment 1, the Stroop effects were larger for comparisons of numerical magnitude than for comparisons of numerosity: for RT, $F(1, 19) = 12.14$, $p < .01$, $MSE = 2{,}891.45$; for error, $F(1, 19) = 6.80$, $p < .05$, $MSE = 80.38$. We can not offer a handy

Table 2
Mean Reaction Times (RTs; in ms) and Percent Errors (PEs) for Comparison of Numerical Magnitude and Numerosity Across Task (Garner Interference) and Across Pair Type (Stroop Congruity) in Experiment 2

	Numerical magnitude		Numerosity	
Task/stimuli	RT	PE	RT	PE
		Garner analyses		
Baseline task	468 (59)	2.47 (5.22)	457 (56)	2.15 (4.71)
Filtering task	487 (74)	4.68 (11.98)	477 (77)	3.33 (8.80)
Garner interference	19**	2.21**	20**	1.18
		Stroop analyses		
Congruent pairs	453 (59)	1.06 (4.36)	456 (66)	1.39 (5.40)
Incongruent pairs	521 (72)	8.29 (15.57)	498 (82)	5.27 (10.88)
Stroop congruity	68**	7.23**	42**	3.88**

Note. Standard deviations appear in parentheses. Error differences refer to percentage point effects.
** $p < .01$.

explanation for this difference in the face of matched dimensional discriminability.[5]

Discussion

Judgments on both dimensions were plagued by Garner interference. Our participants did not ignore irrelevant variation in numerical magnitude when comparing numerosity, and, similarly, they did not ignore irrelevant variation in numerosity when comparing numerical magnitude. Another demonstration of the absence of selective attention was provided by Stroop congruity. Performance on either dimension improved for congruent pairs and deteriorated for incongruent pairs. Our participants ignored neither irrelevant variation nor the momentary content of values along the irrelevant dimension.

In Experiment 1, using binary-valued dimensions, selective attention to numerosity was good; in Experiment 2, using many-valued dimensions, selective attention failed. Set size, hence dimensional uncertainty, was larger in Experiment 2 than in Experiment 1. The change caused the breakdown of selective attention in Experiment 2. Because participants must always pay attention to the relevant dimension to be able to perform the task, of most consequence is the variation of values along the irrelevant dimension. The irrelevant dimension is noticed, and interference ensues.

The results of Experiment 2 demonstrate that matched discriminability (prevailing in Experiment 2) does not suffice to eliminate interference and ensure good selectivity of attention. Matched discriminability is a necessary condition to achieve that goal, but, alone, is not sufficient. Other contextual stipulations must also be satisfied (Algom et al., 1996; Arieh & Algom, 1996, 2002; Bauer & Besner, 1997; Besner & Stolz, 1999; Dishon-Berkovits & Algom, 2000; Francolini & Egeth, 1980; Sabri et al., 2001; Shalev & Algom, 2000).

The results of Experiment 2 can be taken as evidence either to support theories positing the mandatory activation of meaning or to support our account stressing dimensional salience and attention. To decide, we purposely mismatched relative dimensional discriminability favoring numerical magnitude (Experiment 3) or numerosity (Experiment 4). Observing fairly constant interference would be consistent with the standard automaticity account that the processing of meaning is ballistic and obligatory. Altered patterns of interference would be better accounted for by a theory that roots the Stroop effect in attentional shifts wrought by variations of context.

Experiment 3

Method

Participants. The participants were 20 Bar-Ilan University undergraduates, paid volunteers, none of whom had participated in any of the previous experiments. Their ages ranged between 20 and 26 years.

Stimuli and procedure. Apparatus, stimulus values, viewing conditions, and procedure were those of Experiment 2. However, to render

[5] A contentious issue in the literature relates to a distinction between separate processes of assessing numerosity: subitizing and estimating. *Subitizing* refers to the rapid and accurate apprehension of the numerosity of small sets of objects. By contrast, *estimating* (or *counting*) refers to the more laborious and error-prone process of enumerating larger sets of objects. The distinction is mainly based on a discontinuity or "elbow" of the latencies for numerosity judgments around a set size of 3 or 4 (e.g., Kaufman, Lord, Reese, & Volkmann, 1949). Discrepant accounts have been proposed for the boundary separating subitizing and counting (see the reviews by Balakrishnan & Ashby, 1992; Mandler & Shebo, 1982; Trick & Pylyshyn, 1994). Some investigations challenge the very existence of a distinct subitizing process (cf., Balakrishnan & Ashby, 1992). We tested for a subitizing effect in our data, as the numerosities used in our Experiments 2–8 spanned the range of both processes. It should be noted, though, that in our study, the participants were requested to compare numerosity, a task for which exact estimation of numerosity is not necessary. Both RTs and error rates were relatively low for both the smallest (2 and 3) and the largest (7 and 8) values of numerosity, but relatively long for the intermediate values (4, 5, and 6). Thus, our pattern of results does not accord with the monotonically increasing RT function found in studies of the subitizing process. Rather, the obtained function is the bowed "end-anchor" effect, characteristic of comparative judgment (Banks, 1977).

numerosity less discriminable than numerical magnitude, we slightly modified the stimulus displays. We enlarged the filler asterisks to be the same size as that of the numerals (27 × 18 pixels). Because comparisons of numerosity require the detection of all the numerals in each display, we expected it to be impaired by our manipulation. Preliminary testing confirmed our prediction: Comparisons of numerosity turned out to be more difficult than comparisons of numerical magnitude for all the 15 pairs of stimuli used.

Results

The results are summarized in Table 3. Numerical magnitude was compared faster by 73 ms than numerosity at baseline; means of 520 and 593 ms, respectively, $t(19) = 5.85$, $p < .01$. For error, the respective means for numerical magnitude and numerosity were 1.88% and 2.36%, $t(19) = 1.88$, $p > .05$. Overall, we were quite successful in rendering numerical magnitude more discriminable than numerosity.

For numerical magnitude, the more discriminable dimension, Garner interference amounted to a mere 2 ms ($F < 1$). For numerosity, in contrast, Garner interference was 31 ms, $F(1, 16) = 14.70$, $p < .01$, $MSE = 11{,}915.16$. Comparisons of numerosity were disrupted by irrelevant variation in numerical magnitude, but comparisons of numerical magnitude were not affected by variation in numerosity, $F(1, 19) = 8.83$, $p < .01$, $MSE = 5{,}116.56$. For accuracy, the same small amount of Garner interference (0.52%) obtained for both numerical magnitude, $F(1, 16) = 4.75$, $p > .05$, $MSE = 13.67$, and numerosity, $F(1, 16) = 4.01$, $p > .05$, $MSE = 15.85$.

For Stroop congruity, congruent pairs were compared faster than incongruent pairs for both numerical magnitude, $F(1, 16) = 81.96$, $p < .01$, $MSE = 4{,}780.94$, and numerosity, $F(1, 16) = 56.31$, $p < .01$, $MSE = 11{,}290.90$. Congruent pairs were also compared more accurately than incongruent pairs for numerical magnitude, $F(1, 16) = 28.49$, $p < .01$, $MSE = 63.51$, and for numerosity, $F(1, 16) = 13.96$, $p < .01$, $MSE = 189.07$. Numerically, Stroop congruity was larger for numerosity than for numerical magnitude for both RT (by 11 ms) and accuracy (by 0.68%), but the difference was not statistically significant: for RT, $F(1, 19) = 1.04$, $p > .10$, $MSE = 4{,}785.02$; for error, $F < 1$.

Discussion

The results of Experiment 3 differ from those of Experiment 2. In Experiment 2, numerical magnitude and numerosity were equally discriminable, and comparable amounts of Garner interference plagued performance on both dimensions. Stroop congruity was also comparable, although it was a bit larger for numerical magnitude. In Experiment 3, discriminability was mismatched in favor of numerical magnitude, and Garner interference was only obtained for comparisons of numerosity—the less discriminable dimension. Overall, in Experiment 3, numerical magnitude intruded more on numerosity than did numerosity on numerical magnitude. The pliability of the interference supports an explanation in terms of attention: When the irrelevant dimension is more discriminable than the target dimension, attention to the latter is impaired and expressed as interference to performance. On the alternative account of automaticity, the interference to numerosity observed in Experiment 3 is the inescapable outcome of the mandatory activation of magnitude. Experiment 4, in which we made numerosity more discriminable than numerical magnitude, pitted the two explanations against one another.

Experiment 4

Method

Participants. The participants were 16 Bar-Ilan University undergraduates, between the ages of 21 and 26 years, who had not participated in any of the previous experiments. They were paid to perform in the experiment.

Stimuli and procedure. Apparatus, stimulus values, viewing conditions, and procedure were the same as those in Experiment 2. However, to render numerosity more discriminable than numerical magnitude, we reduced the size of the filler asterisks to 1 × 1 pixels (subtending a visual angle of 0.03° in both length and width). As in Experiment 3, our manipulation was expected to influence comparison of numerosity, now facilitating it. Pilot testing confirmed that comparisons of numerosity were

Table 3
Mean Reaction Times (RTs; in ms) and Percent Errors (PEs) for Comparison of Numerical Magnitude and Numerosity Across Task (Garner Interference) and Across Pair Type (Stroop Congruity) in Experiment 3

Task/stimuli	Numerical magnitude RT	Numerical magnitude PE	Numerosity RT	Numerosity PE
		Garner analyses		
Baseline task	520 (82)	1.88 (4.47)	593 (100)	2.36 (4.97)
Filtering task	522 (89)	2.40 (6.69)	624 (112)	2.88 (7.83)
Garner interference	2	0.52	31**	0.52
		Stroop analyses		
Congruent pairs	497 (82)	0.95 (3.56)	594 (103)	1.09 (4.21)
Incongruent pairs	546 (89)	3.85 (8.53)	654 (114)	4.67 (9.93)
Stroop congruity	49**	2.90**	60**	3.58**

Note. Standard deviations appear in parentheses. Error differences refer to percentage point effects.
** $p < .01$.

easier than comparisons of numerical magnitude for all the 15 pairs of stimuli used.

Results

The results are shown in Table 4. Average baseline RT was 418 ms for comparisons of numerosity, faster by 65 ms than the average of 483 ms obtained for comparisons of numerical magnitude, $t(15) = 9.39$, $p < .01$. For error, the respective means were 1.52% and 1.92%, $t(15) = 0.89$, $p > .30$. Overall, numerosity was more discriminable than numerical magnitude at baseline.

Garner interference was 28 ms for comparisons of numerical magnitude, the less discriminable dimension, $F(1, 12) = 16.53$, $p < .01$, $MSE = 6{,}333.24$, and 18 ms for comparisons of numerosity, $F(1, 12) = 25.96$, $p < .01$, $MSE = 1{,}953.33$. Garner interference for errors (1.75%) was only found for comparisons of numerical magnitude, $F(1, 12) = 9.96$, $p < .01$, $MSE = 30.08$; Garner interference for numerosity was 0.74%, $F(1, 12) = 4.17$, $p > .05$, $MSE = 17.70$. The difference between the Garner effects for the two dimensions was not quite significant, $F(1, 19) = 3.86$, $p < .07$, $MSE = 10.55$.

Stroop congruity was 26 ms for numerosity, $F(1, 12) = 46.61$, $p < .01$, $MSE = 2{,}318.48$, but 78 ms for numerical magnitude, $F(1, 12) = 274.38$, $p < .01$, $MSE = 2{,}923.10$, three times the value of the former, $F(1, 15) = 96.38$, $p < .01$, $MSE = 1{,}170.61$. For errors, Stroop congruity was 1.55% for numerosity, $F(1, 12) = 5.51$, $.01 < p < .05$, $MSE = 73.33$, but 5.53% for numerical magnitude, $F(1, 12) = 12.51$, $p < .01$, $MSE = 271.10$. Again, numerosity intruded on judgments of magnitude over three times more than did magnitude on judgments of numerosity, $F(1, 15) = 10.48$, $p < .05$, $MSE = 56.41$.

Discussion

Reversing the dimensional imbalance resulted in a pattern of interference that was the mirror image of that obtained in Experiment 3. Garner interference, measured by error, only afflicted judgments of numerical magnitude. For Stroop effects, comparisons of numerical magnitude were greatly disrupted by conflicting values of numerosity, but comparisons of numerosity were affected less by conflicting values of numerical magnitude. Our participants were unsuccessful in attending selectively to the less discriminable dimension of numerical magnitude, but they were somewhat more successful in attending selectively to the more discriminable dimension of numerosity.

Standard accounts of automatic activation of numerical magnitude (see, Dehaene, 1997, and Pavese & Umilta, 1998, 1999, for recent statements) are inhospitable to such reversals in the pattern of interference. The numerals presented in this experiment were exact physical replicates of those presented in the previous experiment. Nevertheless, the participants were better able to ignore their meaning in the current experiment in which the dimension of numerical magnitude was less salient than in the previous experiment in which numerical magnitude was more salient. We attribute the improvement in selective attention to numerosity to the change in relative dimensional discriminability.

The results of omnibus analyses performed on the data of Experiments 3–4 underscore the critical role of attention in numerical cognition. The three-way interaction of relative discriminability (Experiment 3, Experiment 4), criterial dimension (numerical magnitude, numerosity), and task (baseline, filtering) supports the reversal of Garner interference as a function of dimensional imbalance, $F(1, 34) = 10.16$, $p < .01$, $MSE = 348.71$. The interaction of relative discriminability, target, and pair type (congruent, incongruent) supports the reversal of Stroop congruity as a function of dimensional imbalance, $F(1, 34) = 29.67$, $p < .01$, $MSE = 301.04$. The error data mirrored those for RT; however, due to the high levels of accuracy, only the reversal of Stroop congruity was reliable, $F(1, 34) = 10.00$, $p < .01$, $MSE = 4.81$.

Experiment 5

To further test our hypothesis that numerical interactions are malleable to the point of elimination, we probed the extent to which interference to numerosity from irrelevant numerical magnitude was affected by factors such as practice and motivation. Reisberg et al. (1980) demonstrated (for a single pair of stimuli)

Table 4
Mean Reaction Times (RTs; in ms) and Percent Errors (PEs) for Comparison of Numerical Magnitude and Numerosity Across Task (Garner Interference) and Across Pair Type (Stroop Congruity) in Experiment 4

	Numerical magnitude		Numerosity	
Task/stimuli	RT	PE	RT	PE
		Garner analyses		
Baseline task	483 (56)	1.92 (4.38)	418 (53)	1.52 (4.25)
Filtering task	511 (80)	3.67 (10.85)	436 (65)	2.26 (7.42)
Garner interference	28**	1.75**	18**	0.74
		Stroop analyses		
Congruent pairs	472 (70)	0.90 (4.05)	423 (60)	1.48 (6.76)
Incongruent pairs	550 (69)	6.43 (14.28)	449 (68)	3.03 (7.95)
Stroop congruity	78**	5.53**	26**	1.55*

Note. Standard deviations appear in parentheses. Error differences refer to percentage point effects.
* $p < .05$. ** $p < .01$.

that extended practice could insulate judgments of numerosity of interference from irrelevant numerical magnitude. MacLeod and Dunbar (1988) have shown similar effects of practice on judgments of colors and shapes that were cast into a Stroop-like task. Melara and Mounts (1993) found considerable dilution of Stroop and Garner effects through practice for both words and colors when the two dimensions were matched in discriminability.

In this experiment, we tripled the number of practice trials for comparing numerosity. The activation of semantic information could be further depressed if the participants were sufficiently motivated to ignore it. To augment the motivation of participants to attend solely to numerosity, we promised them a cash bonus proportional to the speed and accuracy of their performance. Reducing or eliminating semantic intrusions would be consistent with the view that semantic processing of numerals is an optional process.

Method

Participants. The participants were 10 Bar-Ilan University undergraduates—paid volunteers—who had not participated in any of the previous experiments. Their ages ranged between 21 and 25 years.

Stimuli and apparatus. The stimulus pairs were those of Experiment 2. We had already established in that experiment that the values on each dimension were equally discriminable at baseline. Apparatus, stimulus presentation, and viewing conditions were those of the previous experiments.

Procedure. The participants were instructed to compare the numerosity of the pairs of stimulus displays. They were asked to press the key corresponding to the location (left or right) of the array containing the larger number of numerals. Each person participated in two 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 perform the experimental task. Half the participants performed the baseline task first, and half performed the filtering task first. In each task, the participants first went through 180 practice trials, three times their number in Experiment 2. Following this practice, they then proceeded to complete the experimental trials. The participants could rest after completing each block of 60 trials. An experimental session consisted of 360 practice trials and 360 experimental trials and lasted about 40 min. Following the completion of both sessions, all of the participants were told that their performance was excellent and were paid the maximal bonus.

Results

In Table 5 we summarize the results. Garner interference amounted to a mere 2 ms ($F < 1$) for RT and to a minuscule 0.38% for error, $F(1, 8) = 2.87$, $p > .10$, $MSE = 5.09$. Stroop congruity was small for both speed (13 ms), $F(1, 8) = 7.88$, $.01 < p < .05$, $MSE = 2{,}314.97$, and accuracy (1.76%), $F(1, 8) = 7.71$, $.01 < p < .05$, $MSE = 40.53$.

Selective attention improved considerably with practice. Within the first experimental session, as Figure 2 illustrates, Stroop congruity decreased from 29 ms in the first block of trials to 6 ms in the last block, $F(2, 18) = 4.01$, $p < .05$, $MSE = 152.50$. Stroop congruity also decreased across the two sessions. In the second session, in particular, the Stroop effects vanished for both RT (9 ms), $F(1, 8) = 3.13$, $p > .10$, $MSE = 1{,}423.74$, and error (1.5%), $F(1, 8) = 2.14$, $p > .10$, $MSE = 52.41$. Clearly, interference from irrelevant numerical magnitude can be eliminated through practice.[6]

Discussion

The results of this experiment show that interference to numerosity can be eliminated through practice and motivation. Garner interference was negligible, and Stroop congruity was small throughout the experiment and vanished with practice. Irrelevant numerical magnitude intruded minimally on comparisons of numerosity. We concluded that activation of semantic information is not mandatory but rather optional, subject to strategic influences.

Consider the current results in tandem with those obtained for numerosity in Experiment 2. We presented the same pairs of stimuli matched for numerical magnitude and numerosity. Baseline RTs were virtually identical in the two experiments (457 ms in Experiment 2, 459 ms in Experiment 5). However, the Garner effect obtained in Experiment 2 vanished in Experiment 5, $F(1, 28) = 3.51$, $.05 < p < .08$, $MSE = 318.52$, and the Stroop congruity obtained in Experiment 2 (42 ms) grew smaller by two thirds in Experiment 5 (13 ms), $F(1, 28) = 13.91$, $p < .01$, $MSE = 196.26$. Motivation and practice did make a difference.

General Discussion

The semantic processing of numerals—retrieval of their magnitude—was not inevitable. The extent of semantic activation depended rather on a few stimulus factors governing attention. One important stimulus factor was dimensional discriminability—the size of the psychological distances separating values on the constituent dimensions. Additional factors included the amount of information embedded in the irrelevant component of the stimulus (determined by the number of values along the irrelevant dimension), the degree of practice performing with the criterial dimension, and motivation. Manipulating these factors, we made numerosity or numerical magnitude the more salient dimension. That dimension, in turn, captured attention and was better able to withstand intrusions from the irrelevant dimension. Conversely, when the irrelevant dimension was more salient than the target dimension, it interfered with performance on the target dimension. Variations in the effectiveness of numerical processing, whether numerical magnitude or numerosity, depended lawfully on the allocation of attention.

The malleability of interference is best demonstrated by examining the collective results of Experiments 2–4. As Figure 3 shows, both Stroop and Garner effects changed systematically in response to changes in relative dimensional discriminability. For numerosity, the effects were largest when irrelevant numerical magnitude was of superior discriminability (Experiment 3). The effects grew smaller when the two dimensions were equally discriminable (Experiment 2), and they diminished further when target numerosity was of superior discriminability (Experiment 4). For numerical magnitude, interference was a mirror image of that for numerosity. The interaction of interference, criterial dimension,

[6] Mean overall RT was 491 ms in the first session and 451 ms in the second session, a 40 ms improvement, $t(9) = 3.98$, $p < .01$. Performance did not differ across blocks within a session (means of 492, 491, and 491 ms for the three blocks of trials in the first session, and means of 452, 450, and 452 ms for the three blocks of trials in the second session). The stabilized overall performance within a session likely derives from the extensive practice that preceded each session.

Table 5
Mean Reaction Times (RTs; in ms) and Percent Errors (PEs) for Comparison of Numerosity Across Task (Garner Interference) and Across Pair Type (Stroop Congruity) in Experiment 5

Task/stimuli	RT	PE
	Garner analyses	
Baseline task	459 (64)	2.58 (5.24)
Filtering task	461 (74)	2.96 (6.59)
Garner interference	2	0.38
	Stroop analyses	
Congruent pairs	454 (72)	2.08 (5.70)
Incongruent pairs	467 (76)	3.84 (7.28)
Stroop congruity	13*	1.76*

Note. Standard deviations appear in parentheses. Error differences refer to percentage point effects.
* $p < .05$.

and discriminability supports the plasticity of numerical processing depicted in Figure 3 for both Garner, $F(2, 53) = 6.22$, $p < .01$, $MSE = 316.78$, and Stroop, $F(2, 53) = 15.67$, $p < .01$, $MSE = 291.54$, effects. A final feature of the data depicted in Figure 3 is the symmetry of interference to the criterial dimension, whether numerical magnitude or numerosity. In the extreme (i.e., in the absence of Stroop and Garner interference for numerosity, such as in Experiments 1 and 5, or in Experiments 2–4 for Garner effects for error), the symmetry may betray true inattentional blindness for numerical magnitude.

According to the automaticity account, an asymmetry in interference is the natural outcome of pitting the obligatory processing of magnitude against the controlled processing of numerosity. The present results show instead symmetrical interference (or its lack thereof) across numerical magnitude and numerosity. The allegedly automatic processing of numerical magnitude was subject to interference from irrelevant numerosity to the same extent that numerosity was vulnerable to irrelevant numerical magnitude.

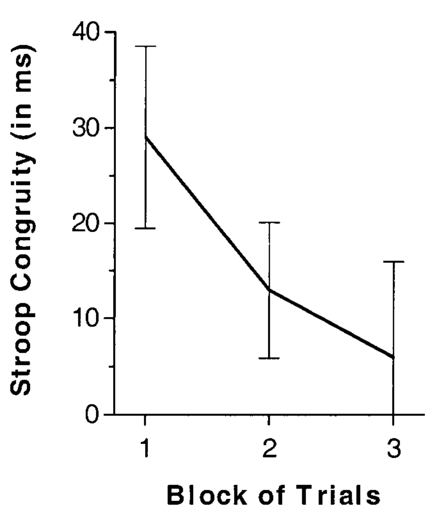

Figure 2. Stroop congruity for comparisons of numerosity for three blocks of trials (Experiment 5, first experimental session). The vertical bars mark one standard error of the mean.

Moreover, Stroop (and Garner) effects were not always present in the pertinent processing. Despite years of education and regardless of mathematical sophistication, people can sometimes treat numerals as mere graphical marks on a computer screen (Algom et al., 1993, 1996; Pomerantz, 1991; Rees, Russell, Frith, & Driver, 1999; Stolz & Besner, 1999; see also, Stevens, 1951).[7]

Does the Absence or Attenuation of the Stroop Effect Imply the Absence or Attenuation of Semantic Processing?

In an important recent study, Rees et al. (1999) measured brain activity with functional magnetic resonance imaging (fMRI) as participants viewed words or nonwords at the center of fixation. Behaviorally, the participants ignored word meaning when unattended, just as we found for numerals. However, if the retrieval of meaning is automatic, then differential activation should be detected even for ignored words. The results showed no differential activation for unattended words (compared with nonwords) in either of the classic language areas or in any other area of the visual cortex. The participants were not blind to the mere presence of letters, but rather to their meaning, "to those properties that distinguish words from random strings of consonants" (Rees et al.,

[7] Dehaene, Bossini, and Giraux (1993) and Dehaene and Akhavein (1995) have reported strong spatial-numerical associations for odd–even, same–different, and parity judgments of numerals—judgments unrelated to numerical magnitude per se. Throughout these tasks, large numbers were responded to more quickly when they appeared on 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 colleagues (Dehaene and Akhavein, 1995; Dehaene et al., 1993) have interpreted these spatial-numerical association of response codes (SNARC) effects as evidence for the mandatory activation of semantic processing—even when magnitude information is irrelevant to the task at hand. We examined number-space associations in each of the first four experiments of the current study and found none to exist for comparisons of numerosity. The absence of a SNARC effect in the current comparisons of numerosity is notable. It provides yet another piece of evidence against the notion of automatic activation of numerical magnitude.

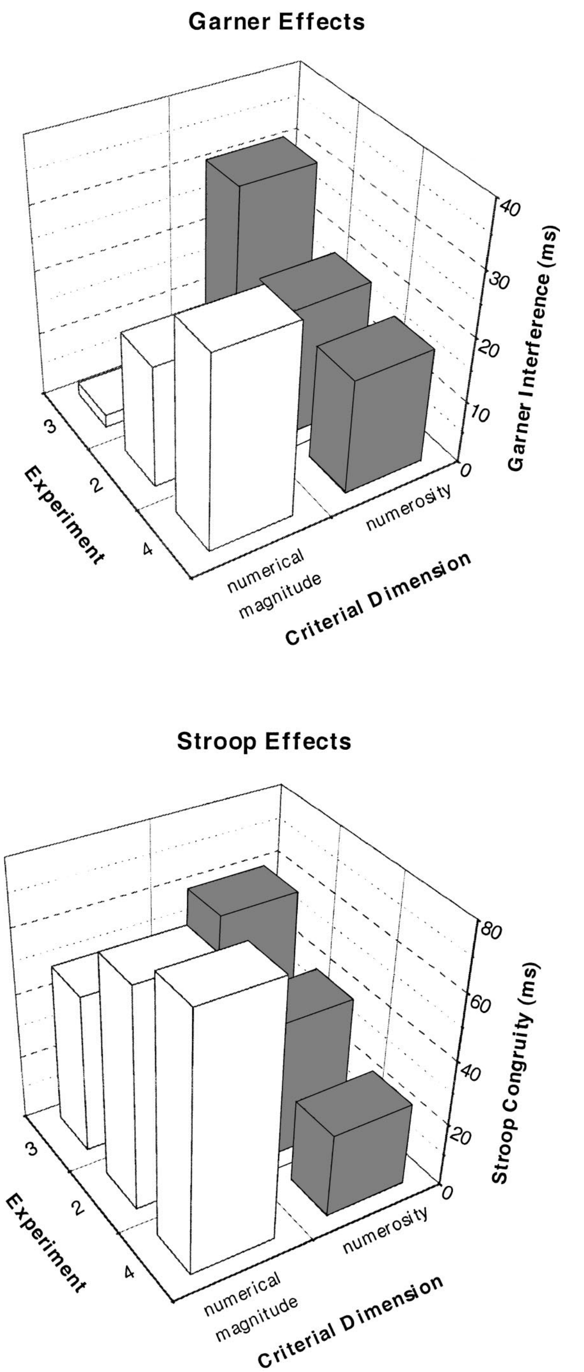

Figure 3. Garner (top) and Stroop (bottom) effects for comparisons of numerical magnitude and numerosity in three experiments with different relative discriminability of the two dimensions. In Experiment 2, the dimensions were equally discriminable. In Experiment 3, numerical magnitude was more discriminable than numerosity. In Experiment 4, numerosity was more discriminable than numerical magnitude.

1999, p. 2507). Commensurate with the current conclusions, Rees et al., stated that "word processing can strongly depend on attention, contrary to previous claims . . . for full automaticity. The data suggest that word processing is not merely modulated but is abolished when attention is fully withdrawn" (p. 2507).

Nevertheless, does the absence or attenuation of Stroop and Garner effects in this study reflect the absence or attenuation of semantic activation? Does the (relative) success at ignoring a bland semantic dimension when irrelevant to the task at hand (i.e., when judging numerosity) reflect true blindness to magnitude or merely sluggish activation of magnitude? On the latter view, the processing of numerosity is sped up such that a decision is reached before magnitude had a chance to interfere. Hence, meaning is activated, but the activation is not expressed by appreciable Stroop and Garner effects. At issue is the interpretability of attenuated or absent effects of Stroop (Besner, 2001; Mari-Beffa, Estevez, & Danziger, 2000; Neely & Kahan, 2001).

One way of uncovering the processing afforded to irrelevant numerical magnitude is examining whether there is a negative priming effect. Negative priming is present in a Stroop-like situation when ignoring one stimulus dimension in a display interferes with responding to the other dimension in a subsequent display when the values are related (cf. Besner, 2001). Negative priming analysis thus provides a means of assessing whether irrelevant magnitude is processed despite the observation of no Stroop effect or of an attenuated one.

We performed analyses of negative priming on the data of Experiments 2–4. We defined a test sequence as one in which the larger value of numerosity (the to-be-produced response) on congruent trial *n* was the value of numerical magnitude (the to-be-ignored response) on incongruent trial $n - 1$. RT performance on trial *n* from test sequences was compared with that measured for the same stimuli from control sequences in which the $n - 1$ trials comprised incongruent stimuli of other experimental values. The mean difference in RT defined negative priming. In Figure 4, we present the effects of negative priming measured in Experiments 2–4.

Figure 4 reveals that negative priming was reliable in Experiment 3, in which numerosity—the target dimension—was less

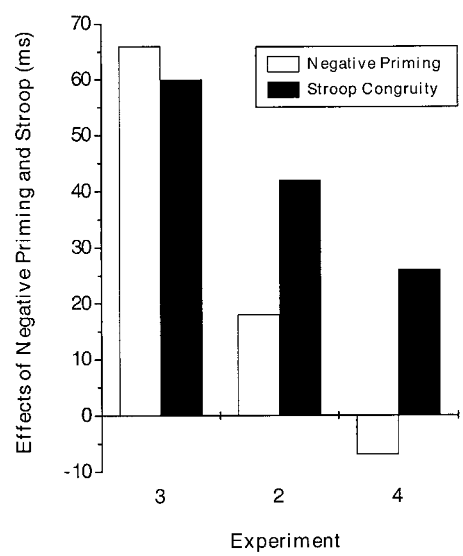

Figure 4. Amount of negative priming and Stroop effects in numerosity performance as a function of relative dimensional discriminability (numerical magnitude more discriminable than numerosity, Experiment 3; numerical magnitude and numerosity equally discriminable, Experiment 2; and numerosity more discriminable than numerical magnitude, Experiment 4).

discriminable than numerical magnitude. In this experiment, negative priming amounted to an appreciable 66 ms, $t(17) = 2.90$, $p < .01$, betraying the extensive processing of irrelevant magnitude. Negative priming was reduced to an insignificant 18 ms, $t(16) > 0.85$, $p > .41$, in Experiment 2 in which the dimensions were equally discriminable, and, notably, the effect disappeared in Experiment 4 (−7 ms), $t(19) = -0.65$, $p > .50$, in which target numerosity was more discriminable than irrelevant magnitude. The interaction of experiment (i.e., relative dimensional discriminability) and negative priming, $F(2, 51) = 4.02$, $p < .05$, $MSE = 3{,}159.00$, documented the fact that negative priming varied reliably with the observed Stroop effect: They increased and diminished in tandem.

The results of the negative priming analysis show that the delayed effects of the distractor—the semantic component of numerals—had no covert influence on responding, certainly not when the distractor was slowed by the discriminability manipulation. The pattern of negative priming faithfully reproduced that of the behavioral Stroop effects, betraying no covert activation of meaning. Incidentally, Besner (2001) and Mari-Beffa et al. (2000) found negative priming in the absence of a Stroop effect. However, as Besner (2001) observed, negative priming per se only shows that some aspect of the number was processed, not necessarily the semantic aspect.

On the basis of other manipulations of context, Stolz and Besner (1999) reached our conclusions, "Stroop interference . . . is not an inevitable consequence of a word's [number's] presentation; words [numbers] do not automatically activate semantic-level representations. Instead, attentional demands modulate the types of analyses performed on the stimulus" (p. 64). We next elucidate the effect on numerical cognition of another variable of context—dimensional uncertainty—then consider the various sources of interference in tasks of comparison.

The Effect of Dimensional Uncertainty

The Stroop effect is traditionally defined as a within-stimulus interference effect that depends on attention to the content (corresponding or conflicting) of the to-be-ignored dimension. Our results, however, show the Stroop effect for a given stimulus to vary as a function of dimensional uncertainty—a factor residing in the entire stimulus ensemble. Consider numerosity performance with the pair of arrays depicted in Figure 1A. The interference recorded for this individual stimulus was larger in Experiments 2–4 than in Experiment 1. The dimensional values were drawn from a larger set in Experiments 2–4 than in Experiment 1, a fact that altered the Stroop effect for the very same stimulus included in the four experiments. Clearly, intrastimulus conflict does not nearly exhaust the sources of interference plaguing performance in a Stroop task. The observation that the processing of the same stimulus varies with the uncertainty context does not accord well with models of automaticity that assume constant processing for a constant stimulus.

The effects of dimensional uncertainty vindicate instead Garner's (1962, 1974) approach to information processing and selectivity. The hallmark of Garner's approach is the contention that performance with a given stimulus depends on the number and identity of the other stimuli that could have appeared on that trial. Our results are consistent with Garner's insights.

Sources of Interference in Classification and Comparison

Originally, Stroop and Garner effects were derived in experiments of classification with individual stimuli presented singly on each of many successive trials. The participants' task was to identify the stimulus on the target dimension, ignoring its value on the irrelevant dimension. Subsequently, the effects have been derived in tasks other than classification, notably in the same–different judgment task (Garner, 1988), and in the comparative judgment task (Besner & Coltheart, 1979; Henik & Tzelgov, 1982). In the literature on numerical cognition, in particular, the overwhelming majority of the studies probing the Stroop effect have applied the comparative judgment task used in this study. Considering Stroop and Garner effects derived from different tasks, one must be careful at identifying their sources. As Garner's (1988) incisive analysis makes clear, the sources may partially differ in classification and comparison. As a rule, the effects are larger in the latter task. The effects may be almost unavoidable in comparison, rendering the small amounts derived in this study all the more impressive.

Why does the difference in outcome due to task exist? Applying Garner's ideas to the present task and stimuli, the dimensions of numerical magnitude and numerosity give rise to responses that are not inherently conflicting when presented individually. People refer to numbers by the adjectives *large* and *small*, and they usually refer to numerosity by the adjectives *many* and *few*. Because the quality of correspondence does not apply to these dimensions, they may turn out to be separable or mildly integral in classification. Yet when these same dimensions are combined in a comparison task, response (in)compatibility is ineluctably produced, because now both dimensions must be either larger or smaller. Classification does not introduce compatibility or conflict more than is inherent in the dimensions or when none had existed before, whereas comparison carries a necessary compatibility relation, because any pair of dimensions must share the "larger" or the "smaller" response. Indeed, there exists evidence (Garner, 1988; Santee & Egeth, 1980; see also, Eriksen, O'Hara, & Eriksen, 1982; St. James & Eriksen, 1991; Thomas, 1996) that dimensions that are found to be separable in classification appear integral in comparison.

Therefore, what is revealing about the results of the current study is how small many of the Stroop and Garner effects were (they averaged 19 ms with numerosity equally discriminable or less discriminable than magnitude). Espousing comparison, we imposed the same response options on both numerical magnitude and numerosity; interference was bound to occur as it indeed did—most of the time. The small effects of Garner and Stroop that were recorded likely reflect the peculiar properties of the comparison task rather than true interaction of numerical magnitude and numerosity.

From a methodological point of view, many intricate issues await resolution when one moves from classification to comparison. First, consider the baseline task in comparison. In Experiment 1, in which each dimension had two values, classification and comparison can become indistinguishable. The participants might have followed the instructions and compared numerosity (in which case irrelevant numerical magnitude was indeed neutral), or they could have classified combinations of numerical magnitude and numerosity (in which case the individual combinations are either congruent or incongruent). We gauged performance for specific stimulus pairs at the baseline tasks and found opposing and statistically insignificant trends for RT and error. Hence, our participants likely compared at baseline, yet the possibility of classification should always be entertained.

Second, in the filtering condition of the comparison task, there is both intratrial and cross-trial variation of the irrelevant component. Both sources may contribute to Garner interference (if there is one), and their separate contributions should be carefully disentangled in future research. Given these new sources of interference, it might be important to devise a new term to cover them: "Comparative Stroop" might be appropriate, alluding to the fact that it is a subtype of the classic Stroop effect.

Third, new insights on the source of the original Garner effect in classification are also relevant to the interpretation of Garner effects in comparison. According to Huettel and Lockhead (1999), the original Garner interference arises due to stimulus sequence: The speed with which dimensional values are classified depends on the stimulus that preceded them. This valuable notion does not seem to provide a full explanation for the present data on comparison. First, we used Stroop-like dimensions. Then, Stroop and Garner interference varied noticeably across Experiments 2–4 for similar stimulus sequences. The effects were also influenced by dimensional uncertainty (Experiment 1 vs. Experiments 2–5), and by practice and motivation (Experiment 5). Several sources conspired to generate Stroop and Garner effects in comparison. Nevertheless, Huettel and Lockhead's insights should be appreciated in future research.

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.

Anderson, J. (1995). *Cognitive psychology and its implications*. New York: Freeman.

Arieh, Y., & Algom, D. (1996). Reading words in Stroop-like tasks: The effect of contingency between semantic and visual components. *Psychologia*, *6*, 7–19.

Arieh, Y., & Algom, D. (2002). Processing picture–word stimuli: The contingent nature of picture and of word superiority. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 221–232.

Ashcraft, M. H. (1994). *Human memory and cognition* (2nd ed.). New York: HarperCollins College Publishers.

Balakrishnan, J. D., & Ashby, F. G. (1992). Subitizing: Magical numbers or mere superstition? *Psychological Research*, *54*, 80–90.

Banks, W. P. (1977). Encoding and processing of symbolic information in comparative judgments. *Psychology of Learning and Motivation*, *11*, 101–159.

Bargh, J. A. (1992). The ecology of automaticity: Toward establishing the conditions needed to produce automatic processing effects. *American Journal of Psychology*, *105*, 181–199.

Bauer, B., & Besner, D. (1997). Processing in the Stroop task: Mental set as a determinant of performance. *Canadian Journal of Experimental Psychology*, *51*, 61–68.

Ben-Artzi, E., & Marks, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics*, *57*, 1151–1162.

Ben-Artzi, E., & Marks, L. E. (1999). Processing linguistic and perceptual dimensions of speech: Interactions in speeded classification. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 579–595.

Besner, D. (2001). The myth of ballistic processing: Evidence from Stroop's paradigm. *Psychonomic Bulletin & Review*, *8*, 324–330.

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

Besner, D., & Stolz, J. A. (1999). What kind of attention modulates the Stroop effect? *Psychonomic Bulletin & Review*, *6*, 99–104.

Besner, D., Stolz, J. A., & Boutilier, C. (1997). The Stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review*, *4*, 221–225.

Bever, T. G., Mehler, J., & Epstein, J. (1968). What children do in spite of what they know. *Science*, *162*, 921–924.

Boronat, C. B., & Logan, G. D. (1997). The role of attention in automatization: Does attention operate at encoding, or retrieval, or both? *Memory & Cognition*, *25*, 36–46.

Brainerd, C. J. (1979). *The origins of the number concept*. New York: Praeger.

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

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

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

Dehaene, S. (1997). *The number sense*. New York: Oxford University Press.

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

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

Dehaene, S., & Cohen, L. (1994). Dissociable mechanisms of subitizing and counting: Neuropsychological evidence from simultanagnosis patients. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 958–975.

Dishon-Berkovits, M., & Algom, D. (2000). The Stroop effect: It is not the robust phenomenon that you have thought it to be. *Memory & Cognition*, *28*, 1437–1449.

Dixon, P. (1978). Numerical comparison processes. *Memory & Cognition*, *6*, 454–461.

Eriksen, C. W., O'Hara, W. P., & Eriksen, B. (1982). Response competition effects in same–different judgments. *Perception & Psychophysics*, *32*, 261–270.

Flowers, J. H., Warner, J. L., & Polansky, M. L. (1979). Response and encoding factors in ignoring irrelevant information. *Memory & Cognition*, *7*, 86–94.

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.

Fraenkel, A., Bar-Hillel, Y., & Levy, A. (1973). *Foundations of set theory*. Amsterdam: North-Holland.

Francolini, C. M., & Egeth, H. E. (1980). On the nonautomaticity of "automatic" activation: Evidence of selective seeing. *Perception & Psychophysics*, *27*, 331–342.

Frege, G. (1980). *The foundations of arithmetic: A logico-mathematical enquiry into the concept of number*. London: Basil Blackwell (Original work published 1884).

Gallistel, C. R. (1989). Animal cognition: The representation of space, time and number. *Annual Review of Psychology*, *40*, 155–189.

Garner, W. R. (1962). *Uncertainty and structure as psychological concepts*. New York: Wiley.

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. (1988). The contribution of information theory to psychology. In W. Hirst (Ed.), *The making of cognitive science: Essays in honor of George A. Miller* (pp. 19–35). New York: Cambridge University Press.

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.

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.

Hofstadter, D. R. (1979). *Godel, Esher, Bach: An eternal golden braid*. New York: Vintage Books.

Huettel, S. A., & Lockhead, G. R. (1999). Range effects of an irrelevant dimension on classification. *Perception & Psychophysics*, *61*, 1624–1645.

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

Husserl, E. (1891). *Philosophie der arithmetik* [Philosophy of arithmetic]. Halle, Germany: C. E. M. Pfeffer.

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.

Kanne, S. M., Balota, D. A., Spieler, D. H., & Faust, M. E. (1998). Explorations of Cohen, Dunbar, and McClelland's (1990) connectionist model of Stroop performance. *Psychological Review*, *105*, 174–187.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology*, *62*, 498–525.

Krueger, L. E. (1972). Perceived numerosity. *Perception & Psychophysics*, *11*, 5–9.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293–323.

Lockhead, G. R. (1966). Effects of dimensional redundancy on visual discrimination. *Journal of Experimental Psychology*, *72*, 95–104.

Lockhead, G. R. (1972). Processing dimensional stimuli: A note. *Psychological Review*, *79*, 410–419.

Lockhead, G. R. (1979). Holistic versus analytic process models: A reply. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 746–755.

Lockhead, G. R., & Pomerantz, J. R. (Eds.). (1991). *The perception of structure: Essays in honor of Wendell R. Garner*. Washington, DC: American Psychological Association.

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

Logan, G. D. (1991). Automaticity in memory. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 347–366). Hillsdale, NJ: Erlbaum.

Logan, G. D. (1992). Attention and preattention in theories of automaticity. *American Journal of Psychology*, *105*, 317–339.

Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition*, *7*, 166–174.

Logan, G. D., & Zbrodoff, N. J. (1998). Stroop-type interference: Congruity effects in color naming with typewritten responses. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 978–992.

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

MacLeod, C. M. (1992). The Stroop task: The "gold standard" of attentional measures. *Journal of Experimental Psychology: General*, *121*, 12–14.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 126–135.

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, *111*, 1–22.

Mari-Beffa, P., Estevez, A. F., & Danziger, S. (2000). Stroop interference and negative priming: Problems with inferences from null results. *Psychonomic Bulletin & Review*, *7*, 499–503.

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 384–394.

Marks, L. E. (1989). On cross-modal similarity: The perceptual structure of pitch, loudness, and brightness. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 586–602.

Mehler, J., & Bever, T. G. (1967, October). Cognitive capacity of very young children. *Science*, *158*, 141–142.

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

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

Melara, R. D., & Mounts, J. R. (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.

Neely, J. H., & Kahan, T. (2001). Is semantic activation automatic? A critical re-evaluation. In H. L. Roediger, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 69–93). Washington, DC: American Psychological Association.

Pansky, A., & Algom, D. (1999). Stroop and Garner effects in comparative judgment of numerals: The role of attention. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 39–58.

Pavese, A., & Umilta, C. (1998). Symbolic distance between numerosity and identity modulates Stroop interference. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1535–1545.

Pavese, A., & Umilta, C. (1999). Further evidence on the effects of symbolic distance on Stroop-like interference. *Psychological Research/Psychologische Forschung*, *62*, 62–71.

Petrusic, W. M., & Baranski, J. V. (1989). Semantic congruity effects in perceptual comparisons. *Perception & Psychophysics*, *45*, 439–452.

Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.

Piaget, J. (1965). *The child's conception of number*. London: Routledge.

Pomerantz, J. R. (1983). Global and local precedence: Selective attention in form and motion perception. *Journal of Experimental Psychology: General*, *112*, 516–540.

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. (1964). Information reduction in the analysis of sequential tasks. *Psychological Review*, *71*, 491–504.

Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Erlbaum.

Posner, M. I., & Snyder, C. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Hillsdale: NJ: Erlbaum.

Pufall, P. B., & Shaw, R. E. (1972). Precocious thoughts on number: The long and the short of it. *Developmental Psychology*, *7*, 62–69.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.

Rees, G., Russell, C., Frith, C. D., & Driver, J. (1999, December). Inattentional blindness versus inattentional amnesia for fixated but ignored words. *Science*, *286*, 2504–2507.

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.

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General*, *126*, 288–311.

Russell, B. (1903). *The principles of mathematics*. Cambridge, England: Cambridge University Press.

Russell, B. (1919). *Introduction to mathematical philosophy*. London: Allen & Unwin.

Sabri, M., Melara, R. D., & Algom, D. (2001). A confluence of contexts: Asymmetric versus global failures of selective attention to Stroop dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 515–537.

Santee, J. L., & Egeth, H. E. (1980). Selective attention in the speeded classification and comparison of multidimensional stimuli. *Perception & Psychophysics*, *28*, 191–204.

Shalev, L., & Algom, D. (2000). Stroop and Garner effects in and out of Posner's beam: Reconciling two conceptions of selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 997–1017.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*, 54–87.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*, 127–190.

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.

St. James, J. D., & Eriksen, C. W. (1991). Response competition produces a "fast same effect" in same–different judgments. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure: Essays in honor of Wendell R. Garner* (pp. 157–168). Washington, DC: American Psychological Association.

Stolz, J. A., & Besner, D. (1999). On the myth of automatic semantic activation in reading. *Current Directions in Psychological Science*, *8*, 61–65.

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.

Thomas, R. D. (1996). Separability and independence of dimensions within the same–different judgment task. *Journal of Mathematical Psychology*, *40*, 318–341.

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, *101*, 80–102.

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review*, *94*, 3–15.

Vos, P. G., van Oeffelen, M. P., Tibosch, H. J., & Allik, J. (1988). Interactions between area and numerosity. *Psychological Research*, *50*, 148–154.

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 February 16, 2000
Revision received August 14, 2001
Accepted September 13, 2001 ■