ELSEVIER

Acta Psychologica 111 (2002) 141–154

acta psychologica

www.elsevier.com/locate/actpsy

Is an ant larger than a lion?

Orly Rubinsten *, Avishai Henik

Department of Behavioral Sciences and Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, 84105 Beer-Sheva, Israel

Received 24 August 2001; received in revised form 19 February 2002; accepted 19 February 2002

Abstract

In order to examine the influence exerted by an irrelevant semantic variable in a comparative judgment task, we employed a Stroop-like paradigm. The stimuli were pairs of animal names that were different in their physical and semantic sizes (e.g., ant lion). Participants were asked to judge which of the two words was larger either in physical or in semantic size. Size congruity effect (i.e., faster reaction times with congruent than with incongruent stimuli) appeared in both semantic and physical judgments. The semantic distance effect (i.e., large semantic distances are processed faster than smaller ones), appeared only when the semantic dimension was relevant to the task. The findings indicate that when a word (animal name) is presented, its meaning is accessed automatically. Part of this meaning (at least with our stimuli) relates to the size of the animal in real life. Processing of meaning of the size of the words is carried out in parallel with the extraction of the physical features of the presented stimuli.

PsycINFO classification: 2340
Keywords: Semantic size; Physical size; Automaticity

1. Introduction

In a mental comparison paradigm, participants typically compare two objects on some underlying magnitude in order to determine their relative order. In such a paradigm it is common to find the *distance effect*, i.e., people more rapidly compare objects that are further apart on the dimension of comparison. For example, we can determine more rapidly the larger of 1 and 9 than the larger of 1 and 3 (Dehaene,

* Corresponding author. Tel.: +972-8-6472086; fax: +972-8-6472932.
E-mail address: orlyr@bgumail.bgu.ac.il (O. Rubinsten).

0001-6918/02/$ - see front matter
PII: S0001-6918(02)00047-1

1989, 1996; Dehaene & Akhavein, 1995; Moyer & Landauer, 1967; Tzelgov, Yehene, Kotler, & Alon, 2000). An additional result in comparison tasks is the *size congruity effect* (Banks & Flora, 1977; Paivio, 1975) manifested in an interaction of semantic and physical magnitudes. One approach for testing the size congruity effect is by using a Stroop-like paradigm. Henik and Tzelgov (1982), for example, presented pairs of Arabic numerals and asked participants to relate to the physical size of the digits and to ignore their numerical value, or judge the numerical value of the digits and ignore their physical size. The two digits could be incongruent (i.e., the numerically larger numeral was physically smaller, e.g., 5 2), congruent (i.e., the numerically larger numeral was also physically larger, e.g., 5 2), or neutral (i.e., only the physical size was different and the digits had the same numerical value, or the other way around, e.g., 5 5 or 5 2, for physical comparisons and numerical comparisons, respectively). The *size congruity effect* is manifested by shorter reaction times (RT) for congruent trials than for incongruent trials (Besner & Coltheart, 1979; Dehaene, 1992; Henik & Tzelgov, 1982; Schwarz & Heinze, 1998; Tzelgov, Meyer, & Henik, 1992). In order to examine the interference or facilitatory components of the size congruity effect, one can compare the neutral stimuli to the incongruent stimuli (the interference component) or to the congruent stimuli (the facilitatory component). In general, the outcome of this Stroop-like paradigm suggests an automatic activation of numerical information. That is, when participants judge the physical sizes of digits, they cannot ignore their numerical values.

Although much research on magnitude comparisons has been done using numerical stimuli, there is considerable evidence that comparable effects are observed with many other types of stimuli as well. For example, the distance effect is not specific to digits, it is also obtained for many other dimensions along which individual objects can be ordered, such as the size of animals (Moyer, 1973; Paivio, 1975), or geographical locations (Maki, 1981). Birnbaum and Jou (1990), for example, asked participants to study associations between names of hypothetical persons and adjectives that described them. Participants then pressed one of two keys to indicate which person in each pair of names was more or less likeable. Response times showed the distance effect.

Paivio (1975) investigated the size congruity effect. Participants were presented with pairs of pictures or printed names of animals and objects differing in rated real-life size (e.g., ant–elephant) and asked to choose the conceptually/semantically larger member of each pair. He found a size congruity effect only when using pictures of different sizes but not when using animal or object names printed in different physical sizes. That is, the physical difference between two pictures had an effect when comparing the semantic dimension (i.e., when the physical information was irrelevant). He suggested that picture comparisons involved visual analog representations in long-term memory while comparison of words involved only the verbal system, not entailing analog representations. Hence, the effects of the irrelevant physical dimension appear for pictures but not for words. Note however that in the replication of Experiment 2 there was a trend toward a size congruity effect for words (and not only for pictures) but this was not significant. The lack of the size congruity effect for words seems to deviate from the effects found with digits. Digits are symbolic like words and still, when two digits are compared along the semantic dimension the irrelevant physical size always affects performance (Henik & Tzelgov, 1982; Schwarz & Heinze, 1998; Tzelgov et al., 1992). In the current study we investigated the appearance of size congruity effect with animal names.

It is possible to examine the size congruity effect with physical comparisons when the semantic information is irrelevant. Within the area of digit comparisons several investigators reported this effect. Accordingly, we wished to examine the hypothesis that comparative judgments for words that are ordered along some conceptual continuum (e.g., animal names) are driven by cognitive mechanisms similar to those used in digit comparisons and hence show similar performance characteristics.

If the semantic and the physical dimensions affect different systems (Paivio, 1975) one would expect to find no size congruity effect, neither in physical comparison of words nor in semantic comparison of words. However, if these two dimensions (of the words) affect the same system or the same stage of processing as in the case of digits (Henik & Tzelgov, 1982; Schwarz & Heinze, 1998) then one would expect a size congruity effect in both cases (or at least in one of them).

In order to examine the influence exerted by an irrelevant variable in a comparative judgment task we employed a Stroop-like paradigm. The stimuli were pairs of animal names, and we manipulated the physical sizes of the printed words and their semantic magnitudes. The semantic comparison was similar to Paivio's (1975) experiment (using only words and not pictures). Our main interest was in the physical comparisons (when the semantic dimension was irrelevant) that were not examined before.

In one session participants were asked to decide which word represents a larger animal (semantic comparison) and in the other session participants were asked to decide which word looks larger (physical comparison). Three types of stimuli were used: congruent (e.g., lion ant), neutral (e.g., lion ant for semantic comparison, lion lion for physical comparison) and incongruent (e.g., lion ant).

2. Method

2.1. Size norms

2.1.1. Participants

Fifty-two university students from Ben-Gurion University of the Negev, participated in the experiment in partial fulfillment of a course requirement.

2.1.2. Stimuli

The experiment involved items drawn from a normative list of animal names scaled for size on the basis of participants' ratings. A total of 39 animal names were selected from Paivio's (1975) size rating norms. The animal names, typed in the Hebrew language in one column, appeared in random order on two pages. Participants were asked to rate the sizes of the animal names given the following instructions: "On the following page you will see names of animals. Please rank them according to their size. Circle one of the digits 1–9, with 1 representing the smallest animal (e.g., worm) and 9 representing the largest animal (e.g., hippopotamus)". Mean responses for each of the 39 words were calculated. The animal names were ranked from smallest to largest in mean rated size, and then divided to three groups containing 13 words each: the largest animals (mean: 8.01, SD 1.034), middle size animals (mean: 4.83, SD 0.96) and the smallest animals (mean: 1.82, SD 0.76). We chose two 4-letter words from each group to create a database from which six stimuli pairs (see Appendix A) were chosen for the experimental block based on semantic distances (described later). We chose three pairs with a semantic distance of 1 and three pairs with a semantic distance of 2. Based on Balgur's (1968) Hebrew language frequency data, there was no significant difference in the language frequency values of the stimuli in Hebrew.

For the practice block, two words from groups 1 and 3, and one word from group 2, with three letters each, were selected to create six pairs of animal names, based on the same logic as that used in the experimental block (see Appendix A).

To control for the possible effect of pre-experimental verbal habits, 52 other participants rated the familiarity of the six experimental and practice pairs. They were asked to rate the familiarity of the words as a pair (i.e., How often did they see or hear the two words together as a pair or in one sentence ?). Participants were asked to rate the familiarity of the pairs on a 9-point scale, were 1 represented an unfamiliar pair and 9, a familiar pair. All the pairs had low means of familiarity (below 4—see Appendix A).

2.2. The experiment

2.2.1. Participants

Sixteen university students (mean age: 22, SD = 2.7) from Ben-Gurion University of the Negev participated in the experiment in partial fulfillment of a course requirement.

2.2.2. Stimuli

A stimulus display consisted of two animal names that appeared centered around a temporary fixation point. The center-to-center distance between two words was 10 mm. Each participant performed two kinds of comparisons. In one, the relevant dimension was physical size, and in the other, semantic value. In every block there were 108 different stimuli that were presented twice (a total of 216 stimuli in every block). Within the set of stimuli prepared for semantic or physical comparison, each word and each physical size appeared on both sides of the visual field an equal number of times. The 108 stimuli included 36 congruent, 36 incongruent and 36 neutral pairs of animal names. A congruent stimulus was defined as a pair of words in which a given word was larger on both the relevant and irrelevant dimensions (e.g., lion–ant). A neutral stimulus was defined as a pair of words that differed only on the relevant dimensions (e.g., lion–ant, for the semantic comparisons or lion–lion, for the physical comparisons). An incongruent stimulus was defined as a pair of words in which a given word was simultaneously larger on one dimension and smaller on the other (e.g., lion–ant). The two words in each pair could be of the same physical size (in which case the pair served as a neutral for semantic comparisons) or could differ in height (physical dimension) by 1 mm (i.e., one word was 6 mm and the other was 7 mm), by 2 mm (i.e., one word was 6 mm and the other was 8 mm) or by 4 mm (i.e., one word was 7 mm and the other was 11 mm). In addition, the two words in each pair could be of the same semantic value (in which case the pair served as a neutral for physical comparisons) or could differ in semantic distance. There were two semantic distances (see Appendix A). Each semantic distance included three different pairs of words. In short, each block had 18 different possible conditions (three physical sizes, two semantic distances and three congruency conditions). Each condition had 12 trials for a total of 216 trials per block.

While the congruent and incongruent trials were the same for the two comparison tasks, the neutral stimuli were different for the two comparisons.

Neutral stimuli in physical comparisons included the same word in two physical sizes. In order to keep the factorial design we created the neutral stimuli from the words that were used for the other two conditions (congruent and incongruent). For example, because the pair "cat–ant" was used to produce congruent and incongruent stimuli for a semantic distance of 1 unit, neutral pairs created by using these two words (e.g., cat cat and ant ant) were included in the analysis as neutral trials for semantic distance 1. Similarly, because the pair "ant–lion" was used for congruent and incongruent conditions for a semantic distance of 2 units, neutral pairs created by using these two words (e.g., ant ant and lion lion) were included in the analysis as neutral trials for semantic distance 2. Each word from all three pairs of a given semantic distance was used in the same number of stimuli. For example, the word "lion" was presented in six stimuli: it was included twice in the analysis as a neutral trial for a physical distance of 1 mm (e.g., lion lion: one word of the pair was 6 mm and the other 7 mm), twice in the analysis as a neutral trial for a physical distance of 2 mm (e.g., lion lion: one word of the pair was 6 mm and the other 8 mm) and twice in the analysis as a neutral trial for a physical distance of 4 mm (e.g., lion lion: one word of the pair was 7 mm and the other 11 mm). Thus, comparisons among congruent, incongruent and neutral conditions were made by using the same words.

Neutral stimuli in semantic comparisons included two words that were different in semantic value but of the same physical size. In order to keep the factorial design we created the neutral stimuli from the words and physical sizes that were used for the other two conditions (congruent and incongruent). For example, the pair "cat–ant" was used to produce congruent and incongruent stimuli for a semantic distance of 1 unit. Hence neutral pairs, created by using these two words (e.g., cat–ant), were included in the analysis as neutral trials for a physical distance of 1, 2 and 4 mm. For a physical distance of 1, the two words (cat–ant) were presented twice in a 6 mm height and twice in a 7 mm height. For a physical distance of 2, the two words (cat–ant) were presented twice in a 6 mm height and twice in a 8 mm height and for a physical distance of 4, the two words were presented twice in a 7 mm height and twice in a 11 mm height. Accordingly, the 7 mm stimuli were divided among the two conditions: half of them were stimuli for a distance of a 1 mm unit and half of them for a distance of a 4 mm unit. In addition, the 6 mm stimuli were divided among the two conditions: half of them were stimuli for a distance of a 1 mm unit and half of them for a distance of a 2 mm unit. Thus, comparisons among congruent, incongruent and neutral conditions were made by using the same physical sizes.

Before every experimental block, participants were presented with a block of 36 practice trials. This block was similar to the experimental block but with the following exceptions: (1) we used different words as stimuli (see Appendix A), and (2) out of the 48 possible congruent trials, we randomly selected 12 trials. The same was done for the 48 possible incongruent and neutral trials.

2.2.3. Design

The manipulated variables were relevant dimension (physical vs. semantic), physical distance (distance of 1, 2 or 4 mm), semantic distance (1 or 2) and congruity (incongruent, neutral or congruent). Thus, we had a $2 \times 3 \times 2 \times 3$ factorial design. All the variables were manipulated within participants.

2.2.4. Procedure

The participant's task was to decide which of two animal names in a given display was larger. Each participant participated in one session composed of two different blocks. In one block, "larger" was defined by physical size and in the other, it was defined by semantic value. The stimuli in each block were presented in a random order. Half of the participants performed physical comparisons first, and the other half performed semantic comparisons first. Before the experiment began, participants were given a practice block. The participants were asked to respond as quickly as possible but to avoid errors. They indicated their choices by pressing one of two keys corresponding to the side of the larger member of the pair.

Each trial began with a fixation point presented for 300 ms. Five hundred ms after the fixation point was eliminated, a pair of animal names appeared and remained in view until the participant pressed a key (but not for more than 5000 ms). A new stimulus appeared 1500 ms after the participant's response.

3. Results

Error rates were generally low and therefore, were not analyzed. In semantic comparisons error rates were 0.21%, 0.28%, and 0.22%, for the congruent, incongruent and neutral conditions, respectively. In the physical comparisons error rates were 0.25%, 0.32%, and 0.15%, for the congruent, incongruent and neutral conditions, respectively.

For every participant in each condition, mean RT for correct trials was calculated. These means were subjected first to a five-way analysis of variance (ANOVA), with relevant dimension (comparison), physical distance, semantic distance, and congruity as within-subject factors and order of tasks (semantic comparison first, semantic comparison second) as a between-subject factor. Order did not produce a significant main effect or an interaction with any of the other variables. Accordingly, we performed a four-way ANOVA, with relevant dimension (comparison), physical distance, semantic distance, and congruity as within-subject factors (with order of tasks omitted).

All four main effects were significant. Physical comparisons were made 230 ms faster than semantic comparisons [$F(1, 15) = 45.19$, MSE = 1692, $p < 0.001$]. Participants responded faster to a large physical distance than to a small physical distance (mean RT for physical distance of 1 mm: 724 ms, for physical distance of 2 mm: 620 ms and for physical distance of 4 mm: 614 ms) [$F(2, 30) = 52.21$, MSE = 14 038, $p < 0.001$]. Similarly, there was the typical semantic distance effect: participants responded faster to a large semantic distance than to a small one (mean RT of semantic distance of 1 unit: 686 ms, 2 units: 620 ms) [$F(1, 15) = 109$, MSE = 5838, $p < 0.001$]. There was a significant size congruity effect [$F(2, 30) = 37.43$, MSE = 9013, $p < 0.001$], with mean RTs of 697, 658 and 609 ms for incongruent, neutral and congruent pairs, respectively.

The *size congruity effect* was dependent on relevant dimension which is indicated by the interaction between relevant dimension × congruity [$F(2, 30) = 3.33$, MSE = 4683, $p < 0.05$], and is presented in Fig. 1.

Since one of the main interests of this experiment was to study the automatic activation of semantic information, we examined the simple effects of congruity for each relevant dimension separately. Congruity was significant both when the relevant dimension was semantic [$F(2, 30) = 29.89$, MSE = 8328, $p < 0.001$] and when it was physical [$F(2, 30) = 19.38$, MSE = 5668, $p < 0.001$]. The size congruity effect was 34 ms larger when the semantic dimension was relevant than when the physical dimension was relevant [$F(1, 15) = 7$, MSE = 4298, $p < 0.01$]. As can be seen in Fig. 1, in the physical comparisons (when semantic dimension was irrelevant),

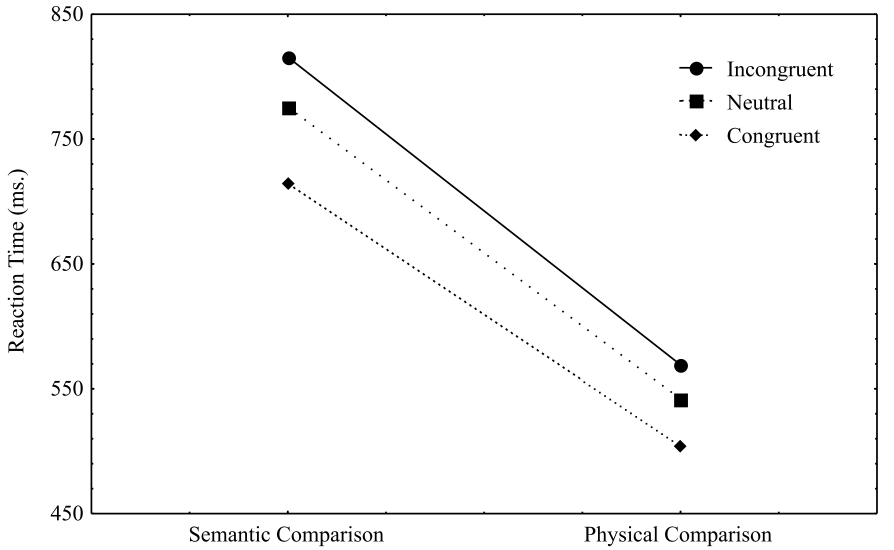

Fig. 1. The size congruity effect: mean RT as a function of relevant dimension (comparison) and congruity.

the size congruity effect was composed of both an interference component (incongruent relative to neutral) [$F(1,15) = 10.95$, $\text{MSE} = 36\,026$, $p < 0.001$] and a facilitatory component (congruent relative to neutral) [$F(1,15) = 12.52$, $\text{MSE} = 69\,847$, $p < 0.001$]. In the semantic comparison (when physical dimension was irrelevant) the size congruity effect was also composed of both an interference component [$F(1,15) = 25.47$, $\text{MSE} = 179\,364$, $p < 0.001$] and a facilitatory component [$F(1,15) = 8.22$, $\text{MSE} = 76\,701$, $p < 0.01$].

The *distance effect* of the semantic dimension was modulated by relevant dimension [$F(1,15) = 75.15$, $\text{MSE} = 11\,168$, $p < 0.001$] and is depicted in Fig. 2. The semantic distance effect appeared in the semantic comparison task (when semantic dimension was relevant) [$F(1,15) = 112$, $\text{MSE} = 13\,084$, $p < 0.001$] but not in the physical comparison task.

It might be argued that the effects we found (i.e., distance and size congruity effect) were due to the fact that there were only six different stimuli and each one of them was presented many times. To examine this possibility, we analyzed the first 12 trials, both in the physical and semantic blocks. The interaction between congruity and task was marginally significant [$F(2,30) = 2.75$, $\text{MSE} = 5079$, $p = 0.07$]. As in the main analysis, we examined the simple effects of congruity for each relevant dimension separately. Congruity was significant both when the relevant dimension was semantic [$F(2,30) = 13.27$, $\text{MSE} = 4675$, $p < 0.01$] and when it was physical [$F(2,30) = 6.26$, $\text{MSE} = 4467$, $p < 0.01$]. In addition, as in the original experiment itself, the distance effect of the semantic dimension was modulated by the relevant dimension [$F(1,15) = 66.42$, $\text{MSE} = 9523$, $p < 0.01$]. The semantic distance effect appeared in the semantic comparison task (when the semantic dimension was rele-

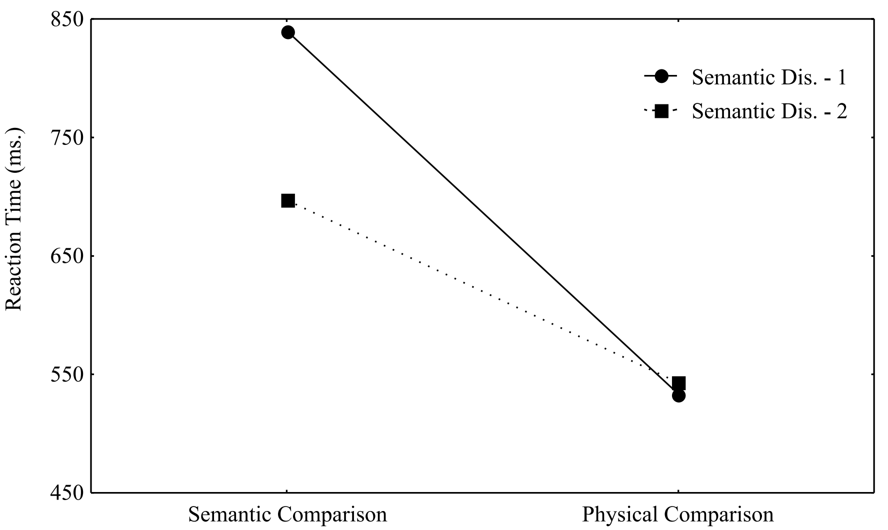

Fig. 2. The distance effect: mean RT as a function of relevant dimension (comparison) and semantic distance.

vant) [$F(1,15) = 30.6$, MSE = 16486, $p < 0.01$] but not in the physical comparison task. Hence, the congruity effect was significant for physical comparisons, meaning that animal sizes were accessed even when irrelevant to the task. Moreover, it should be noted that the present analyses suggest that the reported effects are reliable and robust because statistical significance is achieved even with a small number of trials (only 1–4 trials per condition).

Due to a limit on the number of available words, there was an unbalanced number of repetition of words over the pairings ("elephant" appeared three times, "sheep" appeared one time and all the other four words appeared twice). Hence, one may argue that participants could have developed some sort of strategy. For example, whenever they identified the word "elephant" they did not process the second element of the pairing. This could have happened since "elephant", which occurred much more often, was the largest concept. Hence, participants could have automatically related "elephant" with "large". To test this argument, we analyzed pairs that included animal words that appeared only twice, i.e., ant–cat of semantic distance 1 and fly–lion of semantic distance 2. Each one of the original two blocks contained two trials of these stimuli in each of the 18 conditions (three physical sizes, two semantic distances, and three congruency conditions). Hence, there were four trials per condition. We computed the mean RT for each condition and subjected these means to a four-way ANOVA.

In this analysis, the *size congruity effect* was not dependent on relevant dimension. The interaction between relevant dimension and congruity was not significant ($F < 1$). However, since one of the main interests of this experiment was to study the automatic activation of semantic information, we examined the simple effects of congruity for each relevant dimension separately (as we did in the original analysis). Congruity (incongruent vs. congruent) was significant both when the relevant dimension was semantic [$F(1,14) = 2.7$, MSE = 59968, $p < 0.05$] and when it was physical [$F(1,14) = 5.7$, MSE = 46112, $p < 0.05$]. As in the original analysis, the *distance effect* of the semantic dimension was modulated by relevant dimension [$F(1,14) = 31.6$, MSE = 471, $p < 0.0001$] The semantic distance effect appeared in the semantic comparison task (when semantic dimension was relevant) [$F(1,14) = 6.2$, MSE = 4078, $p < 0.05$] but not in the physical comparison task. Note that aside from the fact that the size congruity effect was similar in size regardless of relevant dimension, the results were similar to those in the original analysis. This suggests that the number of word repetitions did not have any major influence on the pattern of results.

4. Discussion

Let us summarize the main results: (1) Semantic comparisons were slower to judge than physical comparisons. (2) The size congruity effect appeared for both semantic and physical judgments. (3) The semantic distance effect appeared only when the semantic dimension was relevant to the task (i.e., in semantic but not in physical comparisons).

It should be mentioned that, in contrast with our findings, Paivio (1975, second experiment) did not find a size congruity effect for printed animal and object names when the task was similar to our semantic task (i.e., when participants had to judge which of the two words was larger in life). In his replication of that second experiment, he did find that incongruent word stimuli were slower to judge than congruent word stimuli, but this difference was not significant. We suggest that the difference between our semantic comparison and Paivio's results lies in the different stimuli presented in the two experiments. All our word stimuli had the same number of letters (3 in the practice block and 4 in the experimental block). In contrast, Paivio used 117 words containing five or more letters, from a total of 176 words of varying length. Dehaene (1996) examined the difference between Arabic numerals and alphabetically written digits (e.g., three). His event related potentials (ERP) technique and behavioral results pointed to a word length effect: response times for 4-letter words were 19 ms slower than for 3-letter words. This may indicate that using words of different lengths, as in Paivio's (1975) experiment, might have eliminated the size congruity effect. Another possibility is that the variation in Paivio's experiments was larger than in our experiment and that, in turn, reduced the statistical power in his experiments and the chances of getting a significant size congruity effect. Several factors could contribute to such an increase in variability: changes in word length and the large number of different items (both animal and object names).

As mentioned, our main interest was in the effects of the irrelevant semantic dimension (i.e., in physical comparisons), which were never examined before. We found that the irrelevant differences between the stimuli affected performance, regardless of their qualitative nature (physical or semantic). Although the semantic comparisons were slower to judge than physical comparisons, we found that the slower process (semantic) influenced the faster one (physical). This suggests that both of these processes are executed in parallel.

We also suggest that semantic processing of words, which can be ordered along some conceptual continuum, end up with absolute codes for each word (e.g., each word may be labeled as being either large or small). The system is able to note the difference between the two words on both the relevant and the irrelevant dimensions. Hence, size congruity effect appeared in both the physical and semantic comparisons. However, the fact that the semantic distance between the animal names did not affect performance when irrelevant to the task (physical comparison) may indicate that semantic information is automatically computed by the system but only in an unrefined manner.

The present findings may well indicate that participants first convert the animal names to an analog internal representation and later, compare these representations just as they compare numerical or physical representations. It is still, however, impossible to reach a conclusion regarding the nature of these semantic internal representations. It might be that they are localized along an imagined spatial dimension or that they are images of the concepts. A few theories have been proposed concerning the representation involved in imagery. These theories fall into three categories: (1) Theories which suggest that image representations are analog or picture-like (Kosslyn, 1987). (2) Theories which liken image representation to linguistic descriptions

(Pylyshyn, 1981). (3) Theories which suggest that there may exist multiple image representations corresponding to different task demands (Farah, Hammond, Levine, & Calvanio, 1988). The design of our study does not allow us to determine the exact nature of the semantic internal representation of animal words. According to our results, however, it seems that when a word is presented its meaning is accessed automatically. This meaning seems to include some gross characteristics of the object indicated by the word, like gross size (i.e., large vs. small), frequent color, etc. Access of meaning is carried out in parallel with the extraction of the physical features of the stimuli.

The interaction between relevant dimension and congruity (Fig. 1) suggests that the size congruity effect is larger for the semantic comparison (i.e., physical irrelevant) than for the physical comparison (i.e., semantic irrelevant). We ran an additional analysis in which we calculated the proportion of the size congruity effect for each relevant dimension (i.e., for each participant we computed the difference between the results from the incongruent and congruent conditions and divided this difference by the results from the neutral condition). A comparison of these two values shows that there is no difference in size congruity effect between the two relevant dimensions ($t(15) = 0.308, p = 0.75$). Hence, it is possible that the larger size congruity effect in the semantic comparison task could have been the result of slower RTs in semantic comparisons. Note however that there is a size congruity effect regardless of the task (i.e., relevant dimension). That is, the semantic irrelevant dimension affects physical comparisons in spite of physical judgments always being faster than semantic judgments. This violates the "empirical Stroop rule" that suggests that the faster dimension affects the slower one but not the other way around. [1] Hence, our results add to the growing number of works (e.g., Dunbar & MacLeod, 1984) suggesting that a simple horse-race model cannot describe the Stroop effect and possibly similar phenomena.

It was suggested (Dehaene, 1996) that number comparison first arose at 174 ms post onset for Arabic digits and at 190 ms for verbal numerals. In addition, Dehaene (1996) suggested that number comparison involves activation of the parieto–occipito-temporal areas of the brain (see also Cohen, Dehaene, Chochon, Lehericy, & Naccache (2000), for a functional magnetic resonance imagery (fMRI) study). However, word meaning is extracted somewhat later and in different brain areas. According to Abdullaev and Posner (1997), Posner and Abdullaev (1999) word meaning is extracted at around 250–270 ms and involves the frontal areas of the brain. Bentin, Mouchetant-Rostaing, Giard, Echallier, and Pernier (1999) argued that word meaning is extracted at around 450 ms and involves the left fronto-central areas of the brain (see also Petersen, Fox, Posner, & Mintun (1988, 1991) for a positron emission tomography (PET) study; and Gabrieli, Poldrack, & Desmond (1998), for a review). These results suggest that brain areas related to number comparisons and to computations of word meaning are different. Moreover, the time course of these two

[1] We thank the anonymous reviewers for their comments regarding the interaction between relevant dimension and congruity and its possible deviation from the "empirical Stroop rule".

processes also indicates that there is no relationship between them. However, the effects that we found with animal words, i.e., size congruity and distance effects, are similar to those found in the number comparison tasks (Besner & Coltheart, 1979; Dehaene, 1992; Henik & Tzelgov, 1982; Schwarz & Heinze, 1998; Tzelgov et al., 1992). Moreover, in both cases the interaction is with the same variable (i.e., animal name × physical size and numerical value × physical size). These latter behavioral findings might insinuate that there is a relationship between brain mechanisms related to computations of number magnitudes and computations of animal sizes. Clearly, more research is needed in order to clarify the puzzle regarding the relationship between computations of animal sizes hinted at in this experiment and computations of sizes carried out by the parietal mechanism suggested by Dehaene (1996) and others (Cohen, Dehaene, & Verstichel, 1994; Dehaene & Cohen, 1991). [2]

Acknowledgements

This work was supported by a grant to A. Henik from the Israel Science Foundation founded by the Israel Academy of Sciences and Humanities.

Appendix A

Semantic distance	Stimuli of experiment blocks	Familiarity of the pair*	Stimuli of practice blocks	Familiarity of the pair*
2	Fly–lion	3.7	Larva–horse	1.7
2	Fly–elephant (female)	4	Cockroach–horse	4
2	Ant–elephant (female)	3	Cockroach–ox	3
1	Cat–lion	2.5	Dog–ox	2.5
1	Sheep–elephant (female)	2	Larva–dog	2
1	Ant–cat	3.3	Dog–horse	3.3

* 1 = unfamiliar pair; 9 = very familiar pair.

[2] ERP studies commonly produce time parameters and some suggestions regarding brain locations of the presumed generator of the ERP. Because of the inherent limitations of ERP studies with respect to localization of brain areas involved in mental processing, the brain areas mentioned in our discussion were confirmed by PET and fMRI studies (see Petersen et al. (1988, 1991) for a PET study of processing of meaning Gabrieli et al. (1998), for a review; and Cohen et al. (2000) for a fMRI study of number processing) that are designed to identify brain areas involved in mental processing.

References

Abdullaev, Y. G., & Posner, M. I. (1997). Time course of activating brain areas in generating verbal associations. *Psychological Science, 8*(1), 56–59.

Balgur, R. (1968). *The basic word list for elementary schools*. Otsar Hamoreh, 270 p.

Banks, W. P., & Flora, J. (1977). Semantic and perceptual processes in symbolic comparisons. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 278–290.

Bentin, S., Mouchetant-Rostaing, M., Giard, J. F., Echallier, J. F., & Pernier, J. (1999). ERP manifestation of processing printed word at different psycholinguistic levels: time course and scalp distribution. *Journal of Cognitive Neuroscience, 11*(3), 235–260.

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

Birnbaum, M. H., & Jou, J. W. (1990). A theory of comparative response times and "difference" judgments. *Cognitive Psychology, 22*, 184–210.

Cohen, L., Dehaene, S., Chochon, F., Lehericy, S., & Naccache, L. (2000). Language and calculation within the parietal lobe: a combined cognitive, anatomical and fMRI study. *Neuropsychologia, 38*(10), 1426–1440.

Cohen, L., Dehaene, S., & Verstichel, P. (1994). Number words and number non-words: a case of deep dyslexia extending to Arabic numerals. *Brain, 117*, 267–279.

Dehaene, S. (1989). The psychophysics of numerical comparison: a reexamination of apparently incompatible data. *Perception & Psychophysics, 45*, 557–566.

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

Dehaene, S. (1996). The organization of brain in number comparison: event-related potentials and the additive-factor method. *Journal of Cognitive Neuroscience, 8*(1), 47–68.

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., & Cohen, L. (1991). Two mental calculation systems: a case study of severe acalculia with preserved approximation. *Neuropsychologia, 29*, 1045–1074.

Dunbar, K., & MacLeod, C. M. (1984). A horse race of a different color: stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human, Perception and Performance, 10*(5), 622–639.

Farah, M., Hammond, D., Levine, D., & Calvanio, R. (1988). Visual and spatial mental imagery: dissociable systems of representation. *Cognitive Psychology, 20*, 439–462.

Gabrieli, J. D., Poldrack, R. A., & Desmond, J. E. (1998). The role of left prefrontal cortex in language and memory. *Proceedings of the National Academy of Sciences USA, 95*, 906–913.

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.

Kosslyn, S. M. (1987). Seeing and imagining in the cerebral hemispheres: a computational approach. *Psychological Review, 94*, 148–175.

Maki, R. (1981). Categorization and distance effects with spatial linear orders. *Journal of Experimental Psychology Human Learning and Memory, 7*(1), 15–32.

Moyer, R. S. (1973). Comparing objects in memory: evidence suggesting an internal psychology. *Perception & Psychology, 13*, 180–184.

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

Paivio, A. (1975). Perceptual comparisons through the mind's eye. *Memory & Cognition, 3*, 635–647.

Petersen, S. E., Fox, P. T., Posner, M. L., & Mintun, M. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature, 331*(6157), 585–589.

Petersen, S. E., Fox, P. T., Posner, M. L., & Mintun, M. (1991). Positron emission tomographic studies of the processing of single words. *Journal of Cognitive Neuroscience, 2*, 153–170.

Posner, M. I., & Abdullaev, Y. G. (1999). Neuroanatomy, circuitry and plasticity of word reading. *Neuroreport for Rapid Communication of Neuroscience Research, 10*(3), R12–R23.

Pylyshyn, Z. W. (1981). The imagery debate: analog media versus tacit knowledge. In N. Block (Ed.), *Imagery* (pp. 151–206). Cambridge, MA: MIT Press.

Schwarz, W., & Heinze, H. J. (1998). On the interaction of numerical and size information in digit comparison: a behavioral and event-related potential study. *Neuropsychologia, 36*(11), 1167–1179.

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

Tzelgov, J., Yehene, V., Kotler, L., & Alon, A. (2000). Automatic comparisons of artificial digits never compared: learning linear ordering relations. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26*(1), 103–120.