Lost in Translation: On the Language-Specificity of Numerical Representation as Viewed through a Bilingual Prism

by

Susanna Tsivkin

B.A. Psychology
New York University, 1994

SUBMITTED TO THE DEPARTMENT OF BRAIN AND COGNITIVE SCIENCES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN BRAIN AND COGNITIVE SCIENCES
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 1998

Signature of Author: _____
Department of Brain and Cognitive Sciences
May 1, 1998

Certified by: _____
Elizabeth S. Spelke
Professor of Psychology
Thesis Supervisor

Accepted by: _____
Gerald E. Schneider
Professor of Neuroscience
Chairman, Committee for Graduate Students

Lost in Translation: On the Language-Specificity of Numerical Representation as Viewed through a Bilingual Prism

by

Susanna Tsivkin

Submitted to the Department of Brain and Cognitive Sciences
on May 1, 1998 in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Brain and Cognitive Sciences

ABSTRACT

Three experiments investigated the role of a specific language in the human representation of number. Russian-English bilinguals learned facts concerning new numerical operations and arithmetic equations involving large exact or approximate numbers (Experiments 1 and 2), and they learned new geographical and historical material involving large numbers, small numbers, and non-numerical information (Experiment 3). After learning different sets of facts in each of their languages, the learners were tested for knowledge of the facts in both languages. In all three studies, bilinguals retrieved information about exact numbers more effectively in the language of training. In contrast, they retrieved information about approximate numbers and non-numerical material with equal efficiency in the two languages. These findings suggest that specifically human numerical representations (such as the representation, "exactly 11") depend in part on language, whereas the representations humans share with other vertebrates (such as the representation, "approximately 10") do not. Language appears to be involved in the representation of exact, large numbers in a variety of contexts: a finding with implications for theories of language, cognition, and math education.

Thesis Supervisor: Elizabeth S. Spelke
Title: Professor of Psychology

Research over the last few decades provides evidence for representations of number in a variety of nonhuman animals (see Dehaene, 1997, and Gallistel, 1990, for reviews). Untrained primates have been found to represent the exact number of objects in a scene, provided the number is small, and to take account of the effects of additions and subtractions of single objects (Hauser, MacNeilage & Ware, 1996; Uller, Carey & Hauser, 1996).

Trained and untrained birds, fish, and mammals also represent the approximate numerosity of larger sets of items (see Boysen & Capaldi, 1993; Davis & Perusse, 1988; Gallistel, 1990, for reviews). After extensive training, several chimpanzees and one parrot have learned symbols for exact numbers of objects in sets as large as 10 (Matsuzawa, 1985; Boysen & Capaldi, 1993; Pepperberg, 1987), and one chimpanzee has used the symbols to enumerate objects under conditions suggesting a process of addition (Boysen & Berntson, 1989). In all the above cases, the performance of nonhuman animals either equaled or exceeded the performance of human infants tested with no training (see Hauser & Carey, in press; Gallistel & Gelman, 1992, for discussion).

These abilities suggest considerable continuity in number representations over human evolution, but there are striking discontinuities as well. Between ages 2 and 4, human children learn verbal counting. Once counting is mastered, children generalize the counting procedure to larger numbers with no evident upper bound and with no specific training: a feat not seen in any trained animal (Gelman & Gallistel, 1978; Wynn, 1990; cf. Matsuzawa, 1985). School children then learn a set of elementary arithmetic facts and calculation procedures that allow them to perform arithmetic operations on all the numbers they can count (Dehaene, 1997). Finally, children and adults extend their number representations beyond the limits of their counting procedures, using arithmetic operations to pick out fractions, zero, and negative numbers (Gelman, 1991). All these developments distinguish human children from the most highly trained non-human animals.

What is the source of these accomplishments? Some have suggested that humans are endowed with a species-specific system of knowledge of number, and that uniquely human number representations arise as children employ this system to single out numerosities and learn about their interrelations (e.g., Carey & Spelke, 1994; Gelman & Gallistel, 1978). Others have proposed that humans are endowed with the same cognitive systems as are other animals, and that our greater attainment of number knowledge stems from quantitative advantages such as a greater memory capacity or general intelligence (e.g., Putnam, 1975; Johnson, 1997).

We take a different tack: Humans may have the same initial number capacities as other animals but may develop new and distinctive number representations as they grow. This development depends perforce on a system of representation that is indisputably unique to humans: a natural language. The species-specific language faculty, in interaction with one or more species-general number faculties, may underlie the distinctively human capacity for large, exact number.

Connections between knowledge of language and knowledge of number have been suggested on theoretical and empirical grounds (see Dehaene, 1997, for review). Chomsky (1986), noting that the sentences of a language and the numbers in a counting sequence both have the property of discrete infinity, suggested that the same recursive device generates representations of these entities (cf. Bloom, 1994; Hurford, 1987). Neuropsychologists have found that disorders in number representation are frequently accompanied by disorders in language (Warrington, 1982; Dehaene & Cohen, 1991; McCloskey, 1992). Students of cognitive development have observed that advances in abilities to represent numbers accompany the onset of verbal counting (Gelman & Gallistel, 1978; Wynn, 1990) and that efficiency of arithmetic calculation is related to the efficient articulation of number words (Ellis, 1992). Joining them have been students of structural linguistics, positing complementary accounts of representational economy (Becker & Varelas, 1993; Vera & Simon, 1993; see General Discussion below).

Finally, many observers have noted that speakers of two or more languages tend to count and perform arithmetic in just one of their languages--usually, the language in which they originally learned the elementary arithmetic facts. For example, a person who learns arithmetic in one language may move to a different language community, become dominant in the new language, and yet default to the first language when adding up a bill or counting change. According to bilinguals' subjective accounts as well as experimental findings, bilinguals solve arithmetic problems with greater speed and accuracy when problems are presented in the language in which they learned new arithmetic facts (Kolers, 1968; Gonzalez & Kolers, 1987; Marsh & Maki, 1976; Frenck-Mestre & Vaid, 1993; McClain & Huang, 1982).

All the empirical findings, however, can be interpreted in two ways. First, it is possible that numbers and arithmetic facts are represented in the specific natural language in which they are learned. When problems are presented in a different language, they either must be translated to the language of learning or their solutions must be learned anew. The longer response times and lower accuracy at retrieving arithmetic facts in a second language therefore would stem either from a translation process (Marsh & Maki, 1976) or from less well-established, new fact-learning in the second language (Dehaene, 1997). Second, it is possible that numbers and arithmetic facts are represented in a language-independent manner. In order to access those representations, however, one must transform the spoken problem into a representation in the language-independent system in which the answer is computed, and then transform the result of the computation back into the spoken language for production (McCloskey, 1992). These decoding and encoding processes might proceed automatically, even when no spoken response is required, producing the language-specific effects described above.

In the present experiments, we investigated whether different kinds of number facts are represented in language-dependent or language-independent systems, by using a method that distinguishes language-specific encoding and decoding processes from

language-specific representations. We followed the tradition of investigating language and number through studies of bilinguals, with two innovations. First, we conducted bilingual training studies, in which subjects learned new number facts in each of their languages and then were tested on those facts in both languages. This method allowed us to determine whether subjects showed language-specific training effects only for the language in which they habitually performed arithmetic or for their non-preferred language as well. It also allowed us to begin with a clean slate, as it were: We were able to test for language-dependent number representations independently of language-specific encoding and decoding processes, since we could teach subjects different facts about the same numbers in their two languages and therefore give equal training to encoding and decoding processes in the two languages.

Second, the primary question behind our studies is not *whether* subjects show language-specific learning of number facts but *where* they show language-specificity and where they do not. We tested the hypothesis that learning of new facts drawing on humans' unique representations of exact and large numerosities is language-dependent, whereas learning of new facts drawing on the number representations humans share with other animals is not. By this hypothesis, large, exact number facts learned in one language should not be directly accessible to queries in the other language, but non-numerical facts and facts about large, approximate numbers should be equally accessible to queries in both languages.

We began with a study comparing bilingual learning of the exact results of large-number additions and of the approximate results of logarithmic and cube-root operations (Experiment 1). Next we compared bilingual learning of the exact and the approximate sums and products of pairs of large numbers (Experiment 2). Our final experiment compared bilingual learning of large, exact numerical and non-numerical information in fictitious history and geography lessons. After presenting these experiments and their findings, we propose an account of the role of language in number representations.

Experiment 1

Russian-English bilingual college students were trained on two sets of problems involving relations among large, approximate numbers. In one language, they learned the approximate cube roots of each of a set of large numbers; in the other language, they learned the approximate base-two logs of each of a set of large numbers. The students also were trained on four sets of problems involving large, exact numbers. Two sets of problems involved the familiar operation of addition in base ten: adding either 54 or 63 to each of a set of two-digit numbers. Two further sets of problems involved the less-familiar operations of addition in base six and base eight. Each student learned one set of base-ten addition facts and one set of novel-base addition facts in each language, in the same session in which they learned the cube-root or logarithm facts in that language. Half the problems in each set involved specific numbers that appeared only in one language over the course of the experiment, and half the problems involved numbers appeared in both languages, in different problem sets. For the latter problems, exposure to the number words and practice at any decoding and encoding processes involving those words were equated across the two languages.

After two days of training (over which subjects' performance improved for all types of problems), subjects were tested on all the problems in both languages. The speed and accuracy of their responses were compared in the trained and untrained languages, for each set of problems and for both monolingually- and bilingually-presented numbers. If the anecdotes and findings of language-specific calculations stem from processes that translate between the language of input and output and a language-independent system of representation, then performance in the language of training should exceed performance in the untrained language for all problems involving monolingually-trained numbers. If the anecdotes and findings stem from the use of a language-specific system of representation, in contrast, then performance in the language of training should exceed performance in the

untrained language for all problems that require this system, whether the problems involved monolingually- or bilingually-trained numbers. We predicted that problems involving exact, large numbers would show this language-specificity, and that problems involving approximate large numbers would not.

A secondary purpose of this study was to investigate whether training on specific items within a class of problems generalized to other problems within the class, and whether such generalization varied with the language of testing. During the test sessions, subjects were presented with some items on which they had not been trained (e.g., cube-root problems involving new numbers of comparable magnitude to those in the trained problems, and addition problems in which a new number was incremented by 54). Performance on these problems was compared both across the two languages and to performance on the trained problems.

Method

Subjects. Participants were four male and four female bilingual speakers of Russian and English, ranging in age from 18 to 24 years (mean, 19.8 years). All were undergraduate or graduate students living in Ithaca, NY, solicited through the ethnic clubs on the Cornell University campus. To be eligible for participation, the subjects were required to be native speakers of Russian who spoke no English before adolescence, who had spent at least 3 years in the United States, and who were now comfortable conversing in and understanding both Russian and English. The mean age at which the subjects started learning English was 15.4 years (range, 13 to 19 years), and the mean time since coming to the US was 3.8 years (range, 3 to 5 years).

Although all subjects spoke and comprehended Russian and English with ease, subjects were also required to pass a small arithmetic pretest in both languages before participating in the experiment (see below). One additional subject was dropped from the study because the pretest revealed his extremely slow response times for problems given in both languages.

Materials. The experiment was conducted on a Macintosh SE computer. All stimuli appeared on the computer's 30-cm screen with a picture size of 15 cm x 10 cm. The screen was erased before the next stimulus was presented; instant switching was used to mask the previous stimulus. The displays in the two languages were designed to be as similar as possible in size and layout. For the stimuli in both languages, all questions appeared in font size 48, and answers appeared in font size 28. The Geneva font was used for the stimuli in English, and the Russian Central font was used for those in Russian.

Stimuli consisted of six categories of arithmetic problems, described below, with one problem and two potential answers presented on the screen for each trial. All problems and answers were written out in numerical words either in Russian or in English, with the two candidate answers appearing below the problem to the left and right of center (see Table 1 for an example). Subjects selected the correct answer by pressing a key located on the side where that answer appeared ("a" or "k"). The display remained on the computer screen until the subject answered the question, after which a feedback display, specifying whether the answer was correct or incorrect, appeared on the screen for 600 msec. Before the start of each set of training and test items, subjects were given three filler items to accustom or reaccustom them to the problems and procedure.

Insert Table 1 about here

The six categories of problems were as follows:

(1) Addition with addend 54: Subjects were taught to add 54 to each of 12 numbers ranging in value from 47 to 95. The two alternative answers were the correct sum and a distractor in which the tens place differed from the correct answer by one, with the differences balanced in the underestimation and overestimation directions.

(2) Addition with addend 63: Subjects were taught to add 63 to each of 12 numbers ranging in value from 39 to 96. Again, the alternatives were the correct answer and a distractor in which the tens place was raised or lowered by one.

(3) Addition in base 6: Subjects were taught the sums of 12 base-six additions in which the first addend ranged from 1 to 3-3, the second addend ranged from 2 and 5-1, and the sum ranged from 3 to 1-2-1. The response alternatives were the correct answer and a distractor in which either the sixths or the units place was raised or lowered by one from the correct answer.

(4) Addition in base 8: Subjects were taught the sums of 12 base-eight additions in which the first addend ranged from 1 to 4-3, the second addend ranged from 3 to 4-7, and the sum ranged from 4 and 1-1-5. The response alternatives again were the correct answer and a distractor in which either the eighths or the units place was raised or lowered by one from the correct answer.

(5) Estimation of cube roots: Subjects learned to estimate the cube roots of 12 numbers ranging from 9 to 5830. When the correct answer was 4 or less, the distractor differed from it by one in either direction. When the correct answer was larger, the distractor differed from it by 2 in either direction.

(6) Estimation of logs base 2: Subjects learned to estimate the base-two logarithms of 12 numbers ranging from 9 to 8250. When the correct answer was 4 or less, the distractor was one unit larger or smaller. When the correct answer was larger, the distractor was 2 units larger or smaller.

Before the study proper, subjects were given a pretest composed of a variety of items with simple addition, multiplication, division, and subtraction problems in both Russian and English. The problems were presented on the same computer, with all numbers written as in the experiment. Subjects responded to these two-choice questions by pressing one of two keys on the computer, as in the main experiment. Only subjects whose mean reaction times in Russian were under 2500 msec, and in English under 3000

msec, and whose mean reaction times in both languages were within 500 msec of each other were permitted to continue with the experiment. One subject was dropped from the study when the pretest detected his overall deficits in both Russian (mean RT = 2815 msec) and English (mean RT = 3550 msec), as well as his overwhelming superiority in Russian over English (mean difference in RT = 735 msec).

Design. Subjects were given two training sessions and one test session in each of their languages, with Russian and English sessions occurring in alternation. Each training session consisted of three blocked sets of items, with addition of 54, addition in base 6, and cube-root problems occurring during one session and addition of 63, addition in base 8, and log problems occurring during the other session. Each test session consisted of all six blocked sets of items. The order of languages during training, the pairing of languages and problem sets during training, and the order of languages during testing were orthogonally counterbalanced across subjects.

Each set of training problems consisted of 6 repetitions of each of the 12 items in the set, for a total of 72 trials per set (216 trials per session). Each set of test problems consisted of 2 repetitions of each of 6 trained items, and 2 repetitions of each of 6 untrained items, for a total of 24 trials per set (144 trials per session). The correct answer appeared on the left for half the problems in each training and test set. For half the problems, moreover, the numbers in the answers appeared in different problems in the other language ("bilingually trained numbers"). Problems were presented in a random order within each set of 24 items with the restriction that no problem occur twice in succession. Each set of training and test problems was preceded by three filler items.

Procedure. All training sessions in each language were preceded by greetings and casual conversation in that language. Since all subjects lived and worked in an English-speaking environment, a further effort was made to reaccustom them to working with Russian-language materials. Immediately prior to each training session in Russian,

subjects read a different portion of a transcribed lecture in Russian (Brodsky, 1992) and conversed informally about it with the experimenter.

For all training sessions, each problem set began with instructions specific to that problem set and with sample problems, presented in the language appropriate to that session. Throughout training and testing, subjects initiated the first trial by pressing the space bar, and they terminated the trial by pressing a response key indicating whether the number on the left or right correctly answered the problem. Feedback specifying whether the response was correct or incorrect appeared on the screen immediately after the subject's response and remained on the screen for 600 msec. The next trial began immediately after the termination of the feedback display, with the appearance of the next problem on the screen.

Before each testing session, subjects conversed with the experimenter in the language to be used during that session. The procedure for the test trials was identical to that for the training trials, except as follows. Before each block of trials, an announcement appeared on the computer indicating the category of problems subjects were to solve (e.g., addition in base six). No specific instructions or examples were given to the subjects before the test trials began.

Subjects were encouraged to respond efficiently, with equal emphasis on speed and accuracy throughout the training and test sessions. They were given the opportunity to take breaks after each set of 72 items during the training sessions and after each set of 48 items during the testing sessions. Each training session lasted about 45 min, and each test session lasted about one hour.

Results

Training sessions. Table 2 presents the mean reaction times and error rates for each set of problems on each day of training. Accuracy increased and latency declined from day 1 to day 2 for all the types of problems in both languages. A 3 (task) by 2 (language of training: Russian vs. English) by 2 (training day) ANOVA on the response latencies

revealed main effects of training day, F (1, 7) = 183.7, $p < .001$, and task, F (2, 14) = 27.3, $p < .001$, with no interactions. The first effect indicates that subjects responded faster at the end of training, irrespective of the language of training or the task. The second effect was explored further with a Tukey HSD test (with $\alpha = .01$) which indicated that subjects performed slower on the base-ten addition task (mean RT = 3781 msec) than on either the cube root and log tasks (mean RT = 3064 msec) or addition with unfamiliar bases task (mean RT = 3321 msec).[1] Similar to the ANOVA on the response latencies, a 3 (task) by 2 (language of training: Russian vs. English) by 2 (training day) ANOVA on the error rates revealed a main effect of training day, F (1, 7) = 88.12, $p < .001$, with no other effects or interactions being significant.

Insert Table 2 about here

Test sessions. We begin by comparing subjects' performance in the untrained language on problems involving bilingually vs. monolingually trained numbers. A 3 (task: base-ten addition vs. unfamiliar bases addition vs. estimation task) by 2 (language of response: Russian vs. English) by 2 (number training: monolingual vs. bilingual) ANOVA on the response latencies for all test problems presented in the untrained language revealed no main effects or interactions involving number training (all Fs < 3). When bilinguals solved problems in a new language, they were no faster when the numbers in the problems had been trained in both languages (mean RT = 2840) than when the numbers had been trained only in their other language (mean RT = 2741). Moreover, a 3 (task: base-ten addition vs. unfamiliar bases addition vs. estimation task) by 2 (language of response: Russian vs. English) by 2 (number training: monolingual vs. bilingual) ANOVA on error rates demonstrated no main effects or interactions involving number training (all Fs < .3): Error rates did not differ for bilingually versus monolingually trained numbers (respective

means, 1.1% and 1.2%). We therefore disregard the number training factor in all further analyses.

Table 3a presents the mean reaction times and error rates for each set of trained problems when subjects performed in the trained and in the untrained languages. For the tasks requiring exact-number representations, subjects were faster and no less accurate when performing in the language of training, both when that was their first language (Russian) and when it was their second language (English). For tasks involving approximate number representations, in contrast, subjects performed with nearly equal speed and accuracy in the trained and untrained languages.

Insert Table 3 about here

For the latency data, a 3 (task: base-ten addition vs. unfamiliar bases addition vs. estimation task) by 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVA revealed main effects of task, F (2, 14) = 4.7, $p < .05$, and language of testing, F (1, 7) = 49.6, $p < .001$, and an interaction between these factors, F (2, 14) = 17.6, $p < .001$. There was also an interaction between task and language of training, F (2, 14) = 30.9, $p < .001$. A Tukey HSD test (with $\alpha = .05$) applied to the main effect of task revealed that subjects were faster at solving the base-ten addition problems than at solving the addition problems with novel bases. The same Tukey test (with $\alpha = .01$) applied to the interaction between task and language of testing revealed that subjects were faster at solving the base-ten and novel bases addition problems in the language of training, but that they solved the log and cube root problems with equal speed in the two languages. Finally, the Tukey test (with $\alpha = .01$) applied to the interaction between task and language of training revealed that on the base-ten addition problems, subjects performed faster after training in English than after training in Russian, whereas on the addition problems with

novel bases, subjects performed faster after training in Russian than after training in English. For the accuracy data, a 3 (task: base-ten addition vs. unfamiliar bases addition vs. estimation task) by 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVA revealed no significant effects or interactions (all Fs < 2.8).

To understand these results better, separate analyses were performed for each of the three types of task. Starting with the base-ten addition tasks, the 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVA performed on the latency data revealed main effects of language of training, F (1, 7) = 18.9, p < .005, and of language of testing, F (1, 7) = 24.2, p < .005, and no interaction. Subjects responded faster after being trained in English, and they responded faster when queried in the language of training than when queried in the untrained language. A 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVA performed on the error rates revealed no significant effects or interactions (all Fs < 1.1), indicating no speed-accuracy trade-off.

Turning to the addition tasks with novel bases, the 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVA performed on the latency data revealed main effects of the language of training, F (1, 7) = 29.3, p < .005 and of the language of testing, F (1, 7) = 299.8, p < .001, and no interaction. Subjects performed faster after training in Russian, and they performed faster when answering questions in the trained language. Again, error rates did not provide any evidence that a speed-accuracy trade-off contributed to the significant effects found with latency data, as indicated by the 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVA performed on the error rates, which revealed no significant effects or interactions (all Fs < 1.1).

Turning finally to the log and cube root tasks, the 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVA performed on the

latency data revealed no main effects or interactions involving the languages of training or testing (all Fs < 1.2). Similarly, a 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVA performed on the error rates indicated no significant effects or interactions (all Fs < 2.4). Thus, subjects responded with equal speed and accuracy after training in English and Russian, and they responded with equal speed and accuracy when tested in the trained and untrained languages.

Turning now to the question of whether training generalized to new problems, Table 3b presents the mean accuracy and response latencies of performance on the novel problems for each language and task. A comparison of Tables 3a and 3b reveals a difference between the exact and approximate number tasks. For the approximate tasks, untrained and trained problems were solved with equal speed and almost equal accuracy, providing evidence that training generalized beyond the particular problems that subjects learned. For the exact tasks, in contrast, untrained problems were solved slower and with almost equal accuracy than trained problems for both languages of training. These differences were confirmed by 3 (task: base-ten addition vs. unfamiliar bases addition vs. estimation task) by 2 (language of training: Russian vs. English) by 2 (problem novelty: trained vs. untrained problems) ANOVAs on response latencies and error rates, testing problems presented in the language of training. As far as response latencies go, the analysis revealed a main effect of problem novelty, $F(1, 7) = 89.5$, $p < .001$, qualified by interactions between problem novelty and task, $F(2, 14) = 15.6$, $p < .001$, and between problem novelty, task, and language of training, $F(2, 14) = 30.2$, $p < .001$. There was also an interaction between task and language of training, $F(2, 14) = 12.8$, $p < .005$. Tukey HSD tests (all with $\alpha = .01$) applied to the two-way interactions revealed that the subjects were faster at solving old than new problems for both base-ten and unfamiliar bases tasks, but equally fast at solving old and new log and cube root problems. In addition, subjects were faster at solving addition problems in novel bases when trained and tested in Russian, but were equally fast at solving log and cube root problems in the two

languages. A Tukey HSD test (with $\alpha = .01$) applied to the three-way interaction revealed that subjects were faster at solving old than new base-ten addition problems in both languages but with a larger problem novelty effect in English. It also showed that subjects were faster at solving old than new addition problems in unfamiliar bases in both languages but with a larger problem novelty effect in Russian. Finally, the Tukey HSD test revealed that subjects were equally fast at solving old and new log and cube root problems in both languages. With regard to the error rates, please note that ANOVA revealed no significant effects or interactions (all Fs < 2.4).

Subjects' performance on old and new problems presented in the untrained language reveals a further, unanticipated finding: When subjects solved exact number problems (both large number addition and addition in novel bases) in an untrained language, they performed faster and with almost equal accuracy on old problems (i.e., on problems trained in their other language) than on new problems. This effect was confirmed by a 3 (task: base-ten addition vs. unfamiliar bases addition vs. estimation task) by 2 (training language: Russian vs. English) by 2 (problem novelty: old vs. new problem) ANOVA on response latencies to test problems presented in the untrained language. The analysis revealed main effects of task, $F(2, 14) = 22.1$, $p < .001$, and problem novelty, $F(1, 7) = 32.4$, $p < .005$, as well as interactions between task and problem novelty, $F(2, 14) = 10.4$, $p < .005$, language of training and problem novelty, $F(1, 7) = 7.2$, $p < .05$, and task, language of training, and problem novelty, $F(2, 14) = 5.6$, $p < .025$. Tukey HSD tests (with $\alpha = .01$) revealed that when subjects were tested in their untrained language, they were faster at answering log and cube root problems than at answering base-ten or novel-base addition problems, and that they were faster at answering old than new base-ten addition problems. Finally, subjects were faster at answering old than new base-ten addition problems in both languages, and that they were faster at answering old than new addition problems with novel bases when trained in Russian and tested in

English. Subjects solved old and new problems with equal speed when the problems involved logs or cube roots in both languages and when the problems involved addition with novel bases trained in English and tested in Russian.

In probing for accuracy, however, we observed that there were no differences between error rates for old and new problems. A 3 (task: base-ten addition vs. unfamiliar bases addition vs. estimation task) by 2 (training language: Russian vs. English) by 2 (problem novelty: old vs. new problem) ANOVA on error rates revealed no significant effects or interactions (all $\underline{F}$s < .4).

Discussion

Bilingual subjects who learned new facts about approximate numbers in one of their two languages retrieved those facts with equal efficiency in their two languages. This finding qualifies both the anecdotal reports and experimental findings of language-specificity in bilingual arithmetic: When people learn the approximate answers to logarithm and cube-root problems, their learning appears to draw entirely on representations that are language-independent.

In contrast, when the same subjects learned new facts about exact numbers in one of their two languages, they retrieved those facts more efficiently in the language of training than in the untrained language. This finding cannot be attributed to any habitual preference for representing number in one of the two languages, because it was observed for both languages of training: Although all the subjects preferred to perform elementary arithmetic in Russian, new problems taught in English were performed better in English. This finding also cannot be attributed to an effect of training on the efficiency of encoding the number words into a language-independent representation, because subjects performed no better on problems involving bilingually trained number words than on problems involving monolingually trained number words. These findings therefore provide evidence that exact-number facts, involving both the familiar base-ten addition and the more novel addition operation in bases six and eight, are represented at least partly in a language-

specific form, in accord with the theses of Gallistel and Gelman (1992) and Dehaene (1997) and contrary to that of McCloskey (1992).

Students' performance in solving new problems within the training sets (e.g., calculating the cube root of a new 4-digit number, or adding 54 to a new 2-digit number) sheds further light on subjects' learning. For the log and cube-root problems involving approximate number representations, learning generalized fully from old to new problems in both languages. This finding suggests that subjects did not learn a set of individual facts but rather learned to approximate the log and cube-root functions, at least within the range of values in the training set.

For all the exact number problems, two central findings emerged from the analyses of performance on new problems. First, new problems were solved slower and with equal accuracy than old problems. This finding provides evidence that subjects learned individual arithmetic facts during training, rather than a more general procedure for incrementing numbers by a specific amount. Second, subjects showed an advantage for solving old over new problems in the untrained as well as the trained language. In the untrained language, subjects' more efficient performance with old problems might be taken to suggest that the representations underlying exact-number arithmetic learning were not entirely language-specific. As others have proposed (Dehaene, 1997; Gallistel & Gelman, 1992), exact, large number representations may involve both language-dependent and language-independent processes. Alternatively, learning exact-number addition facts may be entirely language-dependent, but subjects may use those facts when queried in a different language by translating the problem into the language of training. In the latter view, subjects may perform more quickly with old than with new problems in the untrained language because translating the old problems into the language of training and retrieving the trained answer is less time-consuming than solving the problems anew in the untrained language. We re-negotiate these two possibilities in Experiment 2.

We have suggested that performance on the log and cube root problems differed from performance on the addition problems because the former problems involved approximate number representations, whereas the latter problems involved exact number representations. Please observe also that the difference between performance with approximate and exact numbers was not due to subjects' unfamiliarity with solving problems involving logs and cubes since both novel bases addition and base 10 addition tasks showed language-specificity.

Nor may the results be attributed to the different levels of difficulty with which unary and binary operations are associated, respectively. A cursory look at our results may lead one to believe that since it might have been easier for subjects to process unary operations than binary ones (and not merely because it may take longer to process lengthier verbal problems), the arguably easier tasks may have been processed language-independently. Recall that during the training phase, subjects did indeed produce answers to logs/cube problems significantly faster than responses to either novel bases or base-ten questions. However, throughout the *testing* phase, subjects tended not to answer logs/cube problems faster than either novel bases or base-ten questions. This possibility is thereby refuted squarely.

Experiment 1 did not address whether it is the kind of arithmetic *task* (e.g., binary vs. unary; addition vs. cube roots) or, alternatively, whether it is the kind of *representation* (exact vs. approximate) that is at the root of language-specificity. Experiment 2 ventures to dissociate these possibilities.

Experiment 2

The primary purpose of Experiment 2 was to compare bilingual students' performance on a single operation--base-ten addition--under conditions in which either an exact or an approximate answer was required. Two groups of students learned the same set of new large-number addition facts in one language (Russian or English) and then were

tested on knowledge of those facts in both languages. Students in one group learned to calculate the exact answers to these facts; students in the other group learned to estimate the approximate answers to these facts. Since the first condition was essentially a replication of the base-ten arithmetic task in Experiment 1, students were expected to perform these problems more effectively in the language of training than in the untrained language. The critical condition was the second: If it is the operation of addition that engenders language-specific processing, then students who learned new approximate-number addition facts in one language should also retrieve the facts more efficiently in the language of training. If, on the other hand, it is the nature of the representation--i.e., exact large numbers--that engenders language-specific processing, then students who learned new approximate number facts in one language should perform equally in their two languages, as did those who learned new log and cube-root estimation facts in Experiment 1.

Experiment 2 had three subsidiary purposes. One purpose was to investigate further the unexpected finding from Experiment 1 that learning to estimate the answers to specific log and cube-root problems generalizes to new problems of the same type, whereas learning exact addition facts does not generalize to new facts and produces an advantage in solving trained over untrained problems in both the trained and the untrained languages. To replicate these findings and determine whether approximate addition training generalizes to new problems, the students in Experiment 2 were tested on new exact and approximate addition problems in each language, and their performance on the new problems was compared to performance on the trained problems.

Another purpose of Experiment 2 was to investigate whether training on the approximate answers to addition problems facilitates performance on tasks requiring calculation of the exact answers to those problems, and vice versa. After learning a set of exact or approximate addition problems, therefore, students in both training conditions were tested on both the exact and the approximate versions of those problems. In accord with the evidence that approximate number representations are automatically activated when

subjects perform exact arithmetic tasks (see Dehaene, 1997; Gallistel & Gelman, 1992, for review), we predicted that training on exact arithmetic problems would benefit performance on the same, approximate addition problems. The existence and extent of transfer in the reverse direction should suggest whether exact number representations are automatically activated during performance of approximate arithmetic tasks.

A third purpose of Experiment 2 was to address a controversy concerning multiplication. Based on his analyses of a variety of animal studies, Gallistel (1990) proposed that the language-independent number representations found in animals serve as inputs to the operation of multiplication. Based on analyses of human patients, in contrast, Dehaene (1997) proposed that approximate number representations enter only into the operations of addition and subtraction. To shed further light on this controversy, the students in Experiment 2 were also taught new multiplication facts under conditions that required either an exact or an approximate answer. If humans have a language-independent system for multiplying approximate large numerosities, then only those subjects taught exact-number multiplication facts should show language-specific learning. If no language-independent multiplication system exists, then all the subjects should show language-specific learning of multiplication facts, regardless of whether exact or approximate answers were required.

Method

Subjects. Participants were 3 female and 5 male bilingual Russian-English speakers, ages 18 to 32 years (mean, 22.5 years). Subjects were solicited through the ethnic clubs at Cornell University, according to the same selection criteria used in Experiment 1. All were undergraduate or graduate students at Cornell who began learning English at a mean age of 15.3 years (range, 12 to 18 years) and who had been in the U.S. for an average of 4.9 years (range, 3.5 to 6.5 years). All subjects spoke and comprehended Russian and English, and passed the pretest from Experiment 1.

Materials. Subjects were tested with the same computer and display set up as in Experiment 1, but with the Adobe Times Ten Cyrillic font used for the Russian stimuli. Four sets of problems were deployed: exact addition, approximate addition, exact multiplication, and approximate multiplication. Each problem was presented as in Experiment 1, with all numbers written in words and with two candidate answers provided (see Table 4 for examples).

Insert Table 4 about here

The problem sets were as follows:

(1) Exact addition. Twelve sums were presented, with the first addend ranging from 22 to 86, the second addend ranging from 18 to 86, and the sum ranging from 47 to 153. The two response alternatives were the correct sum and a number that was ten larger or smaller.

(2) Approximate addition. These problems were identical to the exact addition problems except for the candidate answers, which ranged from 50 to 160. All response alternatives were multiples of ten; one was the nearest such multiple and the other was 30 units larger or smaller than that multiple.

(3) Exact multiplication. Twelve products were presented, with the first factor ranging from 12 to 28, the second factor ranging from 3 to 9, and the candidate answer ranging from 47 to 156. The two response alternatives were the correct product and a number that was ten larger or smaller.

(4) Approximate multiplication. These problems were identical to the exact multiplication problems except for the candidate answers, which ranged from 50 to 160. All response alternatives were multiples of ten; one was the nearest such multiple and the other was 30 units larger or smaller than that multiple.

Design. Each subject participated in four training sessions and two test sessions. A training session presented one set of problems in one language. Each subject learned two sets of problems, one in each language. Half the subjects were trained on exact addition and approximate multiplication problems, and half were trained on approximate addition and exact multiplication problems. The test sessions consisted of blocked presentations of all four sets of problems, with each session in a different language. The pairing of training languages and problems, the order of training sessions, and the order of test sessions were counterbalanced across subjects.

During training, a problem set consisted of 12 different problems presented 6 times (72 problems per set). During testing, a problem set consisted of 6 problems that the subjects were trained on and 6 new problems that the subjects had not encountered previously in the study. Each problem was presented twice during the course of testing (24 problems per set; 96 problems per session). Each blocked set of problems was preceded by 3 filler problems, after which problems appeared in a random order with the restriction that no problem appear twice in succession. Within each set, the correct answer appeared on the left and right with equal frequency.

Procedure. The procedure was the same as for Experiment 1, except as follows. Before the study, both the exact and the approximate tasks were explained, each in the language in which they would be trained. For the approximate calculation problems, subjects were asked not to compute the answer exactly and then choose the answer closest to it, but rather to estimate the answer directly. Since training sessions were short (about 20 minutes), subjects were given two sessions per day, separated by a 10-minute break. Each test session lasted about 45 minutes.

Results: Addition tasks

Training sessions. Table 5 presents the mean response latencies and error rates during training for each of the addition tasks, languages, and sessions. Subjects performed faster and more accurately on the second day of training, showing improvements for both

addition tasks and in both languages. A 2 (task: exact vs. approximate addition) by 2 (language: Russian vs. English) by 2 (session: session 1 vs. session 2) ANOVA on the response latencies revealed only a main effect of session, F (1, 4) = 297.2, $p < .001$, and an interaction of session with task, F (1, 4) = 8.1, $p < .05$. Improvement in performance was greater on the approximate addition task than on the exact addition task. A 2 (task: exact vs. approximate addition) by 2 (language: Russian vs. English) by 2 (session: session 1 vs. session 2) ANOVA on the error rates also revealed a main effect of session, F (1, 4) = 292.7, $p < .001$, and no other significant effects or interactions.

Insert Table 5 about here

Test sessions. Table 6a presents the mean response latencies and error rates for all the trained addition problems, tested in both languages. Subjects who were trained on exact addition problems performed faster and with equal accuracy when they were tested on those problems in the language of training, regardless of whether that language was Russian or English. In contrast, subjects trained on approximate addition problems in either language answered those problems with nearly equal speed and accuracy in the trained and untrained languages. These findings were confirmed by 2 (task: exact vs. approximate calculation) by 2 (training language: Russian vs. English) by 2 (test language: same vs. different) ANOVAs on latencies and error rates. Concerning latencies, the analysis revealed a main effect of the language of testing, F (1, 4) = 19.8, $p < .025$, and an interaction between task and the language of testing, F (1, 4) = 22.0, $p < .01$. Subjects performed better while answering in the language of training for the exact addition problems but not for the approximate addition problems. With regard to error rates, the analysis revealed no main effects or interactions (all Fs < 2.1).

Insert Table 6 about here

Table 6b presents subjects' responses on the new exact and approximate addition problems, both in the language of training and in the untrained language. Comparing Table 6b to 6a, it is clear that subjects trained on approximate addition answered old and new problems with nearly equal speed and accuracy, whereas those trained on exact addition answered old problems more rapidly and with equal accuracy than new problems. These effects were confirmed by a 2 (task: exact vs. approximate addition) by 2 (training language: Russian vs. English) by 2 (problem novelty: trained vs. untrained problem) ANOVAs on the response latencies and error rates on the test problems performed in the language of training. The ANOVA on the response latencies revealed a main effect of problem novelty, F (1, 4) = 9.9, $p < .05$, and a borderline-significant interaction of problem novelty with task, F(1, 4) = 7.7, $p = .051$. Subjects responded faster when solving trained than untrained exact addition problems, but they responded with equal speed when solving trained and untrained approximate addition problems. The ANOVA on the error rates revealed no significant effects or interactions (all Fs < 3.1).

The next analysis compared performance on the old and new problems in the untrained language. First, the latency data was analyzed: A 2 (task: exact vs. approximate addition) by 2 (training language: Russian vs. English) by 2 (problem novelty: trained vs. untrained problem) ANOVA on the test problems performed in the untrained language revealed a main effect of task, F (1, 4) = 11.8, $p < .05$, and an interaction between the task and the problem novelty, F (1, 4) = 7.8, $p < .05$. Approximate addition problems were solved faster, overall, than exact addition problems. Moreover, old exact addition problems were answered faster than new problems in the untrained language, whereas old and new approximate problems were answered with equal speed in the untrained language. Second, the error rates were analyzed: A 2 (task: exact vs. approximate addition) by 2 (training language: Russian vs. English) by 2 (problem novelty: trained vs. untrained

problem) ANOVA on the test problems performed in the untrained language revealed no significant effects or interactions (all Fs < .7).

Finally, Table 7 presents subjects' performance on old arithmetic problems tested in a new format (from exact addition training to approximate addition testing or the reverse). Subjects who were trained on exact addition problems performed equally fast and accurately when they were tested on those problems with approximate and with exact answers required. In contrast, subjects trained on approximate addition problems performed faster and with nearly equal accuracy when tested on those problems for which approximate rather than exact answers were required. This pattern of results held true for subjects trained both in Russian and English. A 2 (task at training: exact vs. approximate calculation) by 2 (training language: Russian vs. English) by 2 (task at test: same vs. different) by 2 (test language: same vs. different) ANOVA on latencies revealed a main effect of the language of testing, F (1, 4) = 14.7, p < .025, and an interaction between the language of testing and task at training, F (1, 4) = 15.8, p < .025. Subjects responded equally fast in the trained and untrained languages after being trained with approximate addition problems, but faster in the trained language after being trained with exact addition problems. Finally, there was a trend toward a significant interaction between the task at training and the task at test, F (1,4) = 5.7, p = .075. The latter result demonstrates that training with exact addition problems improved the subjects' performance at testing on those items for which both exact and approximate answers were required, but training with approximate addition problems facilitated the subjects' performance at testing on those items for which only approximate answers were required.

With respect to error rates, a similar 2 (task at training: exact vs. approximate calculation) by 2 (training language: Russian vs. English) by 2 (task at test: same vs. different) by 2 (test language: same vs. different) ANOVA was performed. The analysis indicated only one significant effect, that of language of testing, F (1, 4) = 8.0, p < .05:

The subjects answered more accurately overall in the language of training as compared to the language in which no training occurred.

Insert Table 7 about here

Results: Multiplication Tasks

Training sessions. Table 8 presents the mean response latencies and error rates for each task (exact vs. approximate multiplication), language, and session. Performance improved from the first to the second session for both tasks and both languages, although performance was faster overall for the approximate problems and the improvement appeared to be steeper for the exact problems. The observations regarding the latency data were confirmed by a 2 (task: exact vs. approximate multiplication) by 2 (language of training: Russian vs. English) by 2 (session: session 1 vs. session 2) ANOVA, which revealed significant main effects of session, F (1, 4) = 164.0, $p < .001$, and task, F (1, 4) = 14.1, $p < .025$, and an interaction of these factors, F (1, 4) = 13.3, $p < .025$. A 2 (task: exact vs. approximate multiplication) by 2 (language of training: Russian vs. English) by 2 (session: session 1 vs. session 2) ANOVA on error rates, however, revealed only one significant result: The subjects performed more accurately during the second session, F (1, 4) = 128.6, $p < .001$.

Insert Table 8 about here

Testing sessions. Table 9a presents the response latencies and error rates for the trained multiplication problems in both the trained and untrained languages. Subjects performed with greater speed and either equal accuracy or a tendency for greater accuracy when they were tested in the language of training, regardless of whether they were trained in Russian or English, and regardless of whether they performed exact or approximate

multiplication tasks. Subjects were also faster but somewhat less accurate at solving the approximate multiplication task than at solving the exact multiplication task. A 2 (task: exact vs. approximate multiplication) by 2 (language of training: English vs. Russian) by 2 (language of testing: same vs. different) ANOVA on response latencies confirmed these observations, revealing significant main effects of language of training, $F(1, 4) = 10.5$, $p < .05$, and task, $F(1, 4) = 13.4$, $p < .025$, and no significant interactions. A 2 (task: exact vs. approximate multiplication) by 2 (language of training: English vs. Russian) by 2 (language of testing: same vs. different) ANOVA on error rates indicated no significant effects or interactions, except for a borderline-significant main effect of task, $F(1, 4) = 6.2$, $p = .07$.

Insert Table 9 about here

Table 9b presents subjects' performance on new multiplication problems of the same type as the trained problems. A comparison of Tables 9b and 9a reveals that subjects performed better on trained than on new problems, regardless of whether the problems involved exact or approximate multiplication. This finding was confirmed by a 2 (task: exact vs. approximate multiplication) by 2 (training language: Russian vs. English) by 2 (problem novelty: old vs. new) ANOVA on response latencies on test problems given in the language of training. The analysis revealed main effects of task, $F(1, 4) = 14.0$, $p < .025$ and problem novelty, $F(1, 4) = 8.3$, $p < .05$, and no interaction between these factors. A 2 (task: exact vs. approximate multiplication) by 2 (training language: Russian vs. English) by 2 (problem novelty: old vs. new) ANOVA on error rates revealed one main effect of task, $F(1, 4) = 10.0$, $p < .05$: The subjects responded less accurately when answering problems involving approximate multiplication.

Finally, subjects performed better on trained than on new problems when tested in the untrained language, indicating some transfer of learning across languages. This effect

was confirmed by a 2 (task: exact vs. approximate multiplication) by 2 (training language: Russian vs. English) by 2 (problem novelty: old vs. new) ANOVA on response latencies on test problems given in the untrained language. Like the previous analysis, this analysis revealed main effects of task, $F(1, 4) = 39.6$, $p < .005$, and problem novelty, $F(1, 4) = 10.1$, $p < .05$, with no interaction. Unlike the previous analysis with error rates on test problems given in the trained language, a 2 (task: exact vs. approximate multiplication) by 2 (training language: Russian vs. English) by 2 (problem novelty: old vs. new) ANOVA on error rates on test problems given in the untrained languages, indicated no significant effects or interactions (all Fs < 6).

Discussion

The findings of Experiment 2 confirm and extend those of Experiment 1. When bilingual subjects learned the exact answers to new large-number addition facts in one of their two languages, they subsequently retrieved those facts more effectively in the language of training than in the untrained language. This replication of Experiment 1 contributes to the evidence for language-specificity in arithmetic learning with large, exact numbers.

In contrast, bilingual subjects who learned the approximate answers to the same large-number addition facts in one language subsequently retrieved those facts with equal speed and accuracy, regardless of the language in which they were queried. This finding provides evidence that facts involving large, approximate numerosities are stored and manipulated in language-independent representations. With Experiment 1, it suggests that the critical difference between language-dependent and language-independent arithmetic tasks stems not from the nature of the operations to be performed but from the nature of the number representations that subjects operate upon: representations of approximate numerosity are language-independent, whereas representations of exact numerosity are language-independent.

Experiment 2 replicated one of the unanticipated findings from Experiment 1: Learning new addition facts involving exact, large numbers did not generalize to other similar facts, but learning new addition facts involving approximate large numbers did generalize to new problems. Like the subjects in Experiment 1 who learned approximate log or cube root facts, those in Experiment 2 who learned approximate addition facts did not appear to store those facts in isolation but rather to construct a more general function that applied to untrained as well as trained problems.

Finally, Experiment 2 supports three further conclusions about language-dependent and language-independent number processing. First, although it provides evidence for a language-independent representation of addition facts for approximate large numbers, it provides no evidence for such a representation of multiplication facts. Training in both exact and approximate multiplication was found to be language-specific, with superior performance in the language of training than in the untrained language. This finding accords with Dehaene's (1997) thesis that nonverbal number representations are accessible to the addition operation but not to multiplication. Because subjects cannot solve multiplication problems through a language-independent representation of approximate numerosities, their only possible strategy on the approximate multiplication tasks is to perform part of the multiplication problem exactly, through the language-dependent process. Performance may be faster overall on the approximate multiplication problems since subjects do not need to complete the full multiplication to choose the closer answer. The underlying processes involved in the exact and approximate multiplication problems, however, appear to be the same (please see the General Discussion to follow for clarification).

Second, performance on exact addition problems tested in the language of training, was better for subjects trained on the exact problems than for subjects trained on approximate versions of those problems. In contrast, performance on approximate addition problems tested in the language of training, was equally efficient regardless of whether

subjects were trained on exact or approximate versions of the problems. This pattern of findings supports Dehaene's (1997) thesis that approximate number representations are automatically invoked during exact calculations, but not the reverse (Dehaene, 1997). If approximate representations are invoked automatically during exact calculation, then a subject who practices an exact calculation will, in effect, receive practice on the approximate calculation as well. This practice may explain the subject's facilitated performance when later, approximate-addition problems are presented for the first time.

The third conclusion concerns the differing effects of training with exact vs. approximate problems on performance in the untrained language. Subjects tested on exact problems in the untrained language performed better when they had been trained in their other language on approximate versions of those problems than when they had been trained on exact versions of the problems, even though the exact versions were identical to the test items (except for the change in language) and the approximate versions were not. This otherwise counterintuitive finding, which is opposite to that obtained for performance in the trained language, further underscores the language-independence of approximate number representations and the language-dependence of exact number representations. When bilingual subjects learn approximate answers to addition problems in one language, they gain representations that are accessible in their second language, fostering performance on both approximate and exact problems in that language. In contrast, when bilingual subjects learn exact answers to addition problems in one language, they appear to gain little language-independent information. As a consequence, performance on exact problems in one language is facilitated more by training in the other language on approximate versions of the problems than by training on the very same exact problems to be solved.

In summary, Experiments 1 and 2 suggest that exact number representations are language-dependent, and approximate number representations are language-independent, across a variety of tasks. Nevertheless, all the tasks tested in these experiments involved arithmetic calculation of some kind. The final experiment investigated whether bilingual

students develop language-dependent representations of large, exact numbers when they learn facts outside the context of any arithmetic problem.

Experiment 3

In our third experiment, we probed for context-dependence. Our primary objective was to tease apart whether it is large exact numerosity alone or large exact numerosity implicated in calculations, that is conceived language-specifically.

In this study, Russian-English bilingual students learned one fictitious history lesson and one fictitious geography lesson over the course of six training sessions. One of the lessons was presented in three sessions in Russian, and the other lesson was presented in the other three sessions in English. Both lessons included material involving large numbers (e.g., "Almost <u>four hundred thirty</u> years ago in a country known as ...") and material involving non-numerical information (e.g., " ... a band of farmers met in a secret, underground <u>cave</u> ..."). The primary purpose of Experiment 3 was to test whether subjects learned numerical and non-numerical material in a language-specific or language-independent manner. To this end, subjects were subsequently tested in one session in Russian and in another in English, on both numerical and non-numerical material in both stories. Subjects' speed and accuracy of response was compared for the language of training and for the untrained language.

A secondary purpose of Experiment 3 was to begin to probe whether other categories of information are stored in a language-specific or language-independent manner. It has been proposed that information about small number fractions (Gallistel & Gelman, 1992), egocentric spatial directions (Hermer & Spelke, 1996), geocentric spatial directions (Levinson, 1996), as well as times of the year and day (Peacocke, 1994) is encoded and retrieved in a language-dependent manner. In an initial effort to explore this possibility, small amounts of information in each of these categories was included in each

story, and small numbers of questions assessing memory for this information were presented during the test, both in the language of training and in the untrained language.

Method

Subjects. Participants were six female and two male bilingual speakers of Russian and English, ranging in age from 19 to 33 years (mean, 24 years). All the retained subjects were undergraduate or graduate students living in the greater Boston, Massachusetts area, solicited through an ethnic club at MIT and posted ads on the campuses of MIT and neighboring universities. Subjects were required to be native speakers of Russian who spoke no English before adolescence, who had spent at least three years in the United States, and who were now comfortable conversing in and understanding both Russian and English. The mean age at which subjects started learning English was 16 years (range, 13 to 18 years), and the mean time since coming to the US was 5 years (range, 3 to 7 years). All subjects spoke and comprehended Russian and English with ease, and passed a comprehension test in each language (see below). One additional subject was dropped from the study for failure to pass this test in English.

Materials. All training was conducted with written stories printed single-spaced on two pages of 8.5 x 11-inch paper, with a count of 3,712 characters, 623 words, and 52 lines for the history story in English; 3,262 characters, 528 words, and 75 lines in Russian; and 4,480 characters, 769 words, and 60 lines for the geography story in English; 3,763 characters, 677 words, and 85 lines in Russian. The stories were printed in English in Times font size 14, and they were printed in Russian in Times Ten Cyrillic font of size 14, equating the presentation of words on a page across the two languages. The history lesson offered an overview of past events in a fictitious country while the geography story detailed the travels and adventures of a fictitious character. In both stories, numbers were written out in numerical words of the appropriate language.

All testing of comprehension and retention took place on a Power PC Macintosh computer with a 43-cm screen. Each question was presented on the monitor with a picture

size of 15 cm x 15 cm. For the training and testing in English, one question appeared on the display in Times font size 48, and two answers appeared below it in Times font size 28. For the stimuli in Russian, questions and answers appeared in Times Ten Cyrillic font in the same point sizes as for the English materials. Each question remained on the computer screen until the subject pressed a key on the side of the correct answer, after which a feedback display appeared on the screen for 600 msec, as in Experiments 1 and 2.

For each of the stories, questions probed subjects' comprehension and memory for information in six categories, as follows.

(1) Exact large number facts. Subjects were queried about 16 large number facts in each story, such as the age of a character, the duration of an event, or the length of a journey. Half of the facts involved numbers that also appeared in the other story in a different context ("bilingually trained numbers"), and half used numbers appearing only in a single story ("monolingually trained numbers"). The large numbers that were tested ranged from 8 to 1993 for the history story and from 6 to 1873 for the geography story. For half the questions presented, the distractor answer was a number that also appeared in the story; for the remaining questions, the distractor answer did not appear in the story. The distractor always differed from the correct answer in either the units or the tens place by any amount between 1 and 9. For half of the questions, the ratio of the correct answer and the distractor was between .89 and .99 (small-split condition). For example, if the correct answer was "430", the distractor was "480." For the other half of the questions, the ratio of the correct answer and the distractor was between .83 and .69 (large-split condition). For example, if the correct answer were "13", the distractor was "18."

(2) Non-numerical facts. Facts were introduced into each story to serve as the basis for 16 questions involving no numbers, whose answers were the names of common objects. For half of the questions in each story, the correct answer and distractor appeared only in a single lesson ("monolingually trained facts"), and for half of the questions the correct answer and distractor also appeared as the two alternative answers for a question in

the other story ("bilingually trained facts"). For example, one question for the history story was, "Where did a band of farmers meet?" with the candidate answers of "in a cave" (correct) and "in a house" (distractor). The corresponding question for the geography story was, "Where did Mary find herself after collapsing suddenly?" with the same candidate answers. For half the questions, the distractors were words drawn from the same story; for the remaining questions, the distractors were words that did not appear in either story.

(3) Exact small number facts. Four facts involving different fractions with numerator and denominator under five (e.g., 1/4, 2/3) and four facts involving the first four ordinal numbers (e.g., first, third) were introduced into both stories and queried for a total of eight questions. All distractors were chosen from the same set of fractions and ordinal numbers as the correct answers.

(4) Spatial facts. Material was introduced into each story to serve as the basis for 12 questions probing the retention of information concerning spatial directions. Eight questions probed the learning of egocentric directions, with the candidate answers of "left," "right," "back," "front," "beside," "behind," "above," and"below". The other four questions probed the learning of geocentric directions, with the candidate answers of "west," "east," "south," and "north." For all the questions, distractors were drawn from the same set of spatial terms as the correct answers. All candidate answers appeared in both stories.

(5) Temporal facts. Material was introduced into each story to serve as the basis for four questions concerning time of day (e.g., mid-morning) and four questions concerning season of the year (e.g., winter). For the time of day questions, distractor terms did not appear in either story; for the season questions, distractors appeared in the same story, and all terms were queried in both stories.

(6) Proper names. Six questions concerning the names of characters and landmarks were used as fillers for each story. None of the queried names appeared in more than one story; all distractors were names that appeared in the same story.

Before the study, subjects were given a two-part comprehension test in both Russian and English. Subjects were first required to name 61 pictures drawn from the categories of common fruits and vegetables, spatial terms, animals, precious stones, and miscellaneous items. Subjects next completed a written questionnaire consisting of simple short-answer questions probing their knowledge of spatial terms (e.g., "what is the opposite of 'near'?"), of the names of times and seasons (e.g., "April occurs in what season?"), and miscellaneous items. Included in the picture-naming task and the written questionnaire were all the terms that would serve as answers and distractors during the training and test sessions. In order not to focus subjects' attention on the terms to be tested, tests of the critical terms were interspersed with a variety of other items. All subjects (except the one who was dropped for failure to name the precious stones in English) correctly named all the key target items in both languages.

Design. Each subject was given three training sessions and one test session in each of their languages, with Russian and English sessions occurring in alternation. The order of languages and the pairing of languages with stories were orthogonally counterbalanced across sessions. During each session, subjects read one story and answered 12 different questions. Over the course of the training sessions, subjects encountered a total of 36 different questions per story, of which 8 tested large number knowledge, 8 tested non-numerical knowledge, 4 tested small number knowledge, 6 tested spatial knowledge, 4 tested temporal knowledge, and 6 were fillers. The sequence and types of questions presented during a particular training session were predetermined by the experimenter and fixed for all subjects. The only randomization allowed was the presentation order of first and second questions after each reading of the story (e.g., for one subject the order might have been question 1, then question 2; and for another subject, question 2, then question 1) in order to preclude a potential order effect.

After training, subjects received one test session in each language. Each test session consisted of 30 questions from the history story and 30 questions from the geography

story. Each set of 30 questions consisted of 8 questions testing large number knowledge, 8 questions testing non-numerical knowledge, 4 questions testing small number knowledge, 6 questions testing spatial knowledge, and 4 questions testing temporal knowledge. Half the questions from each category of questions for each story had been presented during training ("old questions") and half had not been presented previously. Each question was asked twice, for a total of 120 questions/session.

Procedure. At the start of each training session, subjects were given general instructions in the language to be used for that session, after which they were presented with one story. The story was read twice aloud by the subject, twice aloud by the experimenter, and twice silently by the subject, for a total of six readings/session. Each reading of the story was followed by two questions, which were drawn in a predetermined order from the six categories, for a total of 12 questions during the session. Questions were presented on the computer, following the same procedure as for Experiments 1 and 2. Breaks were taken after the first three readings of the story. At the start of each test session, subjects were told, again in the language appropriate for that session, that they would be reading and answering questions about both of the stories that they had studied. Subjects did not read the stories during the testing session but were instead administered the 30 test questions in random order, followed by a repeat of the same 30 questions in a different random order. Breaks were taken after every set of 15 questions.

Results

Training sessions. Table 10 presents the mean reaction times and error rates for questions on stories presented in both languages and on each of the days of training. Accuracy increased and latency declined from day 1 to day 3 for the questions in both languages. Two separate 2 (story) by 2 (language of training) by 3 (training day) ANOVAs on the response latencies and on the error rates were performed. The analysis on response latencies revealed only a main effect of training, F (2, 6) = 107.9, $p < .001$. The effect was explored further with the Tukey HSD test (with $\alpha = .01$), which indicated that

the subjects performed faster on day 3 (mean RT = 2752 msec) as compared to both day 1 (mean RT = 4522 msec), and day 2 (mean RT = 3550 msec), and that they also responded faster on day 2 as compared to day 1. Analogously, the analysis on the error rates indicated a significance of the main effect of training, F (2, 6) = 48.7, $p < .001$. Again, the effect was explored further with the Tukey HSD test (with $\alpha = .01$), which revealed that the subjects responded more accurately on day 3 (mean error rate = 1%) as compared to both day 1 (mean error rate = 19%) and day 2 (mean error rate = 10%), and that they responded with more accuracy on day 2 as compared to day 1.

Insert Table 10 about here

Testing sessions. First we consider subjects' memory for large-number facts and non-numerical material. Table 11a presents the mean reaction times and error rates during the test sessions for large number facts and for non-numerical facts that were tested during training, both in the language of training and in the untrained language. For the large number facts, subjects answered more rapidly and either as accurately or with a tendency for more accuracy in the language of training than in the untrained language, regardless of whether the training language was Russian or English. For the non-numerical facts, subjects answered with equal speed and nearly equal accuracy in the language of training and in the untrained language.[2]

Insert Table 11 about here

We performed two separate 2 (fact type: numerical vs. non-numerical) by 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVAs on latencies and error rates. The analysis on error rates indicated no significant effects or interactions (all Fs < 1.8). The analysis on response latencies, on the other hand,

revealed that the subjects were faster at answering in the trained language than in the untrained language, F (1, 7) = 76.7, $p < .001$, and that this difference was larger after subjects were trained in Russian, as demonstrated by the interaction of the language of training with the language of testing, F (1, 7) = 5.8, $p < .05$. Furthermore, the analysis showed the main effect of fact type, F (1, 7) = 154.1, $p < .001$. This main effect, however, was qualified by two interactions. First, and most importantly, there was an interaction of fact type with language of testing, F (1, 7) = 16.6, $p < .01$: The subjects answered numerical, but not non-numerical, questions faster in the trained than in the untrained language. Second, there was a three-way interaction between fact type, language of training, and language of testing, F (1, 7) = 5.7, $p < .05$: When tested with numerical questions, the subjects responded faster in the trained than in the untrained language after being trained in Russian or English; these differences, however, were larger after subjects were trained in Russian. As for the non-numerical questions, the subjects responded equally fast in both languages after being trained in Russian, and slightly slower in the trained language after being trained in English.

Table 11b presents the mean response latencies and error rates for large number facts and for non-numerical facts that were not tested during training. These measures show the same pattern as in Table 11a: subjects who were queried about large number facts responded faster and with equal accuracy in the language of training, whereas subjects who were queried about non-numerical information responded with equal speed and accuracy in the two languages.

Again, two separate 2 (fact type: numerical vs. non-numerical) by 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVAs on response latencies and error rates were performed. The analysis on error rates revealed no significant effects or interactions (all Fs < 4.3).

The analysis on response latencies revealed that the subjects were faster at answering in the trained language than in the untrained language, F (1, 7) = 108.8, $p <$

.001, and that this difference was larger after subjects were trained in Russian, as demonstrated by the interaction of the language of training with the language of testing, F (1, 7) =12.2, $p < .025$. Furthermore, the analysis showed the main effect of fact type, F (1, 7) = 145.4, $p < .001$: The non-numerical questions were answered faster than those that were numerical. Finally, the main effect of fact type was qualified by an interaction of fact type with the language of testing, F (1, 7) = 13.4, $p < .01$: The subjects answered numerical, but not non-numerical, questions faster in the trained than in the untrained language.

A comparison of Tables 11a and 11b reveals that subjects were nearly equally fast and accurate at answering questions presented during training and at answering new questions for both numerical and non-numerical facts. These findings suggest that subjects learned the stories as wholes, rather than memorizing the answers to specific questions. Again, two separate 2 (fact type: numerical vs. non-numerical) by 2 (language of training: Russian vs. English) by 2 (type of question: old vs. new) ANOVAs on response latencies and error rates with items tested in the language of training were performed. The analysis on error rates revealed two main effects: the language of training, F (1, 7) = 7.0, $p < .05$, and the fact type, F (1,7) = 7.0, $p < .05$. The first effect indicated that the subjects made significantly more errors in Russian than in English, and the second effect demonstrates that the subjects were less accurate while answering the numerical than non-numerical questions. The analysis on response latencies revealed one main effect of language of training, F (1, 7) = 6.2, $p < .05$: Contrary to the subjects' better performance in terms of accuracy after being trained in English, the subjects performed faster after being trained in Russian than after being trained in English. Importantly, there were no main effects or interactions involving the question type variable in either ANOVA on response latencies or error rates.

We then ran two separate 2 (fact type: numerical vs. non-numerical) by 2 (language of training: Russian vs. English) by 2 (question type: old vs. new) ANOVAs on response

latencies and error rates with items tested in the language in which no training occurred. The analysis on error rates indicated no significant effects or interactions (all Fs < 1.2). The analysis on response latencies revealed one significant result: the main effect of fact type, F (1, 7) = 137.0, p < .001: Non-numerical questions were answered faster than numerical questions.

Next we consider subjects' memory for material involving small fractions and ordinal numbers. Table 12a presents the mean response latencies and error rates for "old" questions (i.e., questions presented during training) concerning information in each of these categories. For both questions involving small fractions and small ordinal numbers, subjects were faster and equally or slightly more accurate when tested in Russian than in English, regardless of whether the language of training was Russian or English. Two separate 2 (fact type: small fractions vs. small ordinal numbers) by 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVAs on response latencies and error rates were performed. The analysis on response latencies confirmed these observations, and revealed one significant result: The subjects were faster at answering in the trained language than in the untrained language, but only if the training occured in Russian; having been trained in English, the subjects answered slightly slower in English than in Russian, as indicated by the interaction of the language of training with the language of testing, F (1, 7) =15.0, p < .01. The analysis on error rates, on the other hand, indicated no significant effects or interactions (all Fs < 4.3).

Insert Table 12 about here

Table 12b presents the mean response latencies and error rates for small number questions, separately for small fractions and small ordinal numbers that were not tested during training. This Table shows a pattern similar to that in Table 12a: subjects responded faster in Russian, regardless of whether the training was in Russian or English.

Unlike subjects' responses to the "old" questions, the pattern of accuracy rates with "new" questions (i.e., questions not asked during training) indicated that the subjects were either equally accurate or a bit more accurate in the language of training, regardless of whether the training occurred in Russian or English. Two separate 2 (fact type: small fractions vs. small ordinal numbers) by 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVAs on latencies and error rates, including questions not presented during training, confirmed these observations. As for the analysis on error rates, no significant effects or interactions were found (all Fs < 3.6). As for the analysis on response latencies, the subjects were faster when responding in the trained language, F (1, 7) = 28.3, $p < .005$. This effect was further explained by the interaction of the language of training and the language of testing, F (1, 7) = 10.3, $p < .025$: The subjects were faster in the trained language if that language were Russian, and slightly slower in the trained language if that language were English.

Comparing Table 12b to 12a, one could see that the subjects were equally fast and nearly equally accurate when responding to the questions presented during training and when responding to the new questions, for both small fractions and small ordinal number facts. This pattern of results indicates that the subjects were trained on the content of the stories and not just specific questions. To confirm this observation, we ran two separate 2 (fact type: small fractions vs. small ordinal facts) by 2 (language of training: Russian vs. English) by 2 (type of question: old vs. new) ANOVAs on response latencies and error rates, including questions tested in the trained language. The analysis on error rates indicated no significant main effects or interactions (all Fs < 4.3). The analysis on response latencies revealed no main effect or interactions of type of question or fact type. The only significant effect was of language of training, F (1, 7) = 16.0, $p < .01$. The subjects performed faster after they were trained in Russian than after they were trained in English.

We then ran two separate 2 (fact type: small fractions vs. small ordinal numbers) by 2 (language of training: Russian vs. English) by 2 (type of question: old vs. new) ANOVAs on response latencies and error rates, including questions tested in the new (untrained language). Both analyses on response latencies and error rates revealed no main effects or interactions (all $\underline{F}$s < 3.5).

Finally, we consider subjects' memory for material involving spatial and temporal facts. Table 13a presents the mean response latencies and error rates for "old" questions (i.e., questions presented during training) concerning information in each of these categories, separately for egocentric and geocentric spatial directions, and times of day and seasons. For the questions involving egocentric spatial directions, and times of day, subjects were faster when tested in Russian than in English, regardless of whether the language of training was Russian or English.

Subjects were slightly more accurate in responding to the questions involving egocentric spatial directions in the trained language, regardless of whether the language of training was Russian or English, and equally accurate answering in the trained and untrained languages in responding to the questions involving times of day. For geocentric facts and for facts concerning the seasons, subjects answered faster in Russian when they were trained in Russian; they answered with nearly equal speed in the two languages when they were trained in English. Subjects were equally accurate in responding to the questions probing their knowledge of seasonal facts while answering in the trained and untrained languages. When the subjects responded to the geocentric questions, they were equally accurate answering in the trained and untrained languages when the language of training was Russian, and a bit less accurate in the language of training when that language was English.

Insert Table 13 about here

To explore these results, we first ran two separate 2 (fact type: temporal vs. spatial) by 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVA on latencies and error rates. The analysis on latencies revealed that the subjects were faster when responding in the trained language, F (1, 7) = 16.6, $p < .01$. This effect was further explained by the interaction of the language of training and the language of testing, F (1, 7) = 6.9, $p < .05$: The subjects were faster in Russian, regardless of whether the language was training was Russian or English. The analysis on error rates revealed no significant effects or interactions (all Fs < 2.2)

To explore these results further, we ran separate ANOVAs for each fact type (i.e., temporal and spatial). Two separate 2 (fact type: egocentric vs. geocentric) by 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVAs on latencies and error rates were performed first. The analysis on latencies indicated one significant result: The subjects performed faster in the trained than in the untrained language, F (1, 7) = 16.4, $p < .01$. The analysis on error rates, on the other hand, indicated that none of the main effects or interactions were significant (all Fs < 1.2). Next, another two separate 2 (fact type: seasons vs. time of day) by 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVAs on latencies and error rates were performed. The analysis on latencies revealed an interaction between the language of training and the language of testing, F (1, 7) = 7.1, $p < .05$: The subjects responded faster in Russian, regardless of whether the language of training was Russian or English. The analysis on error rates revealed no significant effects or interactions (all Fs < 1.1).

Table 13b presents the response latencies and error rates for new questions in the same categories. This Table shows a pattern similar to that in Table 13a: subjects responded faster and somewhat more accurately in Russian, regardless of whether the training was in Russian or English for times of day questions. In addition, the subjects responded faster in the language of training after being trained in either Russian or English

with questions concerning seasons, although the subjects had a tendency to be more accurate in the untrained language after being trained in Russian or English. Unlike subjects' responses to the "old" questions, the pattern of responses to the geocentric questions indicated that the responses were faster in Russian, regardless of whether the training occured in Russian or English. The accuracy rates, however, indicated that the subjects had a tendency to be more accurate in the language of training, for both training in Russian and English. Finally, when it came to egocentric questions, the subjects responded faster in the language of training after being trained in Russian, and equally fast in the language of training and testing after being trained in English. Notice, however, that the subjects appeared to be more accurate in English after being trained in either English or Russian. To explore these results, we ran two separate 2 (fact type: temporal vs. spatial) by 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVAs on latencies and error rates. Similar to the analysis of "old" questions, the analysis on latencies with "new" questions revealed that the subjects were faster when responding in the trained language, F (1, 7) = 6.6, $p < .05$. This effect was further explained by the interaction of the language of training and the language of testing, F (1, 7) = 7.2, $p < .05$: The subjects were faster in Russian, regardless of whether the language of training was Russian or English. The analysis on error rates revealed no significant effects or interactions (all Fs < 4.5).

As with the analysis on the "old" questions, we then ran separate ANOVAs for each fact type (i.e., temporal and spatial). Two separate 2 (fact type: egocentric vs. geocentric) by 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVAs on latencies and error rates were performed. The analysis on latencies indicated that the subjects were faster when responding to the questions presented in the trained language, F (1, 7) = 8.9, $p < .025$. This effect, however, was further explained by the interaction of the language of training and the language of testing, F (1, 7) = 6.7, $p < .05$: The subjects responded faster in Russian, regardless of whether they were

trained in Russian or English. The analysis on error rates revealed that none of the main effects or interactions were significant (all Fs < 1.1).

Next, we ran another two separate 2 (fact type: seasons vs. time of day) by 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVAs on latencies and error rates. The analysis on response latencies revealed a three-way interaction between the fact type, language of training and language of testing, F (1, 7) = 7.8, $p < .05$: As discussed above, the subjects responded faster in the language of training when answering questions about seasons, regardless of whether the training occured in Russian or English. On the other hand, the subjects responded faster in Russian when answering questions about time of day, regardless of whether the language of training was Russian or English. The analysis on error rates, again, indicated no significant main effects or interactions (all Fs < .4)

Comparing Table 13b to 13a, one could see the subjects were nearly equally fast and accurate when responding to the questions presented during training and when responding to the new questions for both spatial and temporal facts. This pattern of results indicates that subjects learned the stories as a whole rather than their specific components during training. To confirm these observations, we ran two separate 2 (fact type: temporal vs. spatial) by 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVAs on latencies and error rates. The analysis on latencies indicated that there were no significant main effects or interactions (all Fs < 4.8). The analysis on error rates revealed only one significant result: The subjects answered questions involving spatial facts more accurately than the questions involving temporal facts, as indicated by the main effect of fact type, F (1, 7) = 10.7, $p < .025$.

We then ran separate ANOVAs for each fact type (i.e., temporal and spatial). Two separate 2 (fact type: egocentric vs. geocentric) by 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVAs on latencies and error rates were performed. Both the analyses on latencies and error rates, respectively, revealed

no significant main effects or interactions (all Fs < 2.4). Next, we ran another two separate 2 (fact type: seasons vs. time of day) by 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVAs on latencies and error rates. Again, both analyses on latencies and error rates, respectively, revealed no significant effects or interactions (all Fs < 5.5).

Finally, we ran two separate 2 (fact type: temporal vs. spatial) by 2 (language of training: Russian vs. English) by 2 (questions type: old vs. new) ANOVAs on latencies and error rates with questions presented in the new (untrained) language included in the analysis. Both analyses indicated no main effects or interactions (all Fs < 5). We then ran separate ANOVAs for each fact type (i.e., temporal and spatial). Two separate 2 (fact type: egocentric vs. geocentric) by 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVAs on latencies and error rates were performed. Both the analysis on latencies and the analysis on error rates revealed no significant main effects or interactions (all Fs < 3.4). Next, we ran another two separate 2 (fact type: seasons vs. time of day) by 2 (language of training: Russian vs. English) by 2 (language of testing: same vs. different) ANOVAs on latencies and error rates. Both analyses on latencies and error rates, respectively, revealed no main effects or interactions (all Fs < 4).

Discussion.

Bilingual subjects who learned new numerical material in the context of a fictitious history or geography lesson subsequently retrieved the material with greater speed and accuracy in the language of training than in the untrained language. In contrast, non-numerical material learned during the same lessons was subsequently retrieved with equal ease in the two languages. This pattern of results was detected with both facts tested during the training sessions, and facts not tested during the training, suggesting that the subjects learned the numerical and non-numerical material, and not just the specific answers, in a language-specific and language-independent way, respectively.

These findings provide the first systematic evidence, we believe, that large number representations are language-specific even in contexts having nothing to do with mathematics. Together with Experiments 1 and 2, these findings suggest that representations of exact, large numbers have a language-dependent component in multiple contexts of learning.

The results of Experiment 3 highlight a surprising finding: Our subjects revealed language-dependent performance on numerical material, regardless of whether the ratio of the correct answer and the distractor was small or large. We expected that the subjects would show a language-dependent pattern of results only when the ratio of the two answers was high (e.g., "57" vs "58"), and would behave in a language-independent fashion when the ratio of the two answers was low (e.g., "13" vs "18"). These results were predicted based on animals' ability to set apart numbers of low ratio (e.g., "7" vs. "9") (Gallistel, 1990) but not of high ratio ("7" vs "8"). An explanation of these unexpected results may lie in the strategy subjects selected. If a subject noticed that to answer at least some of the numerical questions she had to remember the number exactly (e.g., exactly "57", and not just "about 60"), she might have chosen to answer all the questions--regardless of whether only the approximate representation might have been adequate--in a language-specific way.

If this explanation is valid, it should be possible to manipulate numerical questions with a low ratio of the correct answer to the distractor, into being answered in a more language-independent way. For example, if the story is set up so that only non-numerical material appears in the story along with large-split numerical material, subjects should perform more language-independently (just as animals do).

Since our subjects were able to answer all numerical questions in the untrained language, it is clear that some learning transferred to the untrained language. Since the lessons contained fictitious material, it did not make sense to ask questions about material not presented in the study; therefore, we could not say how much of the learning transfer

might have been due to either direct translation from the trained into the untrained language or approximate number representation playing a role in subjects' performance. Studies comparing the transfer of exact and approximate number information, like those in Experiment 2, might dissociate these two possible explanations.

Experiment 3 also provides suggestive evidence for a different pattern of language specificity for bilingual subjects' retention of information about fractions involving small numbers, small ordinal numbers, spatial directions, and temporal relationships. Subjects appeared to answer questions with greater speed and accuracy when tested in Russian than when tested in English, especially (but not exclusively) when training was in Russian. This finding may stem from subjects' tendency to encode and remember spatial, temporal, and small-number information in Russian, regardless of whether the information is presented in Russian or English. Since fewer questions were administered in these categories, however, these effects are weaker than the effects concerning large numbers and non-numerical information. More research is needed to investigate why this pattern of language-specificity was observed.

General Discussion.

Jerry Fodor (1975) wrote almost 25 years ago that at the intersection of science and philosophy, speculative psychologists would construct theories of mind. This said, Fodor (1975) set out to sketch an unabashed speculative thesis on the role of language and thought. Likewise, and at the risk of invoking the same sorts of fallibilities, we will articulate a theory of mind, language, and number that will explain both the results of the three experiments presented in this paper, and the work that has preceded it.

The present experiments provide the first clear evidence, from behavioral studies, that language mediates the processing of large, exact numbers. Evidence for language-specificity was observed both when bilingual students learned arithmetic facts involving the familiar operations of base-ten addition and multiplication, when they learned facts

involving the less familiar operations of addition in bases 6 and 8, and when they studied fictitious history and geography lessons.

Language-specificity occurred not only when the students learned new facts in the language in which they habitually perform arithmetic, but also when they learned new facts in their other language: in the latter case, subjects were faster and more accurate at retrieving the facts in a language that they rarely use for representing number, but in which these facts were presented. All these findings contrast, however, with subjects' retrieval of facts about approximate numerosities and about non-numerical information.

When subjects learned new addition facts, logarithms, and cube roots concerning approximate numerosity, and when they learned new historical or geographical material with no numerical, temporal, or spatial component, their retrieval was independent of the language in which the facts were presented. These findings of language-independence support two suggestions about the language-dependent representations described above. First, language-specificity is a property of the learning of large, exact number facts (embedded or not in a story), not of bilingual learning in general since only these facts, as opposed to others that were tested, were language-specific. Second, learning is language-independent when numerosities are approximate and language-dependent when numerosities are exact.

Previous findings indicate that human infants and adults, as well as animals make use of two distinct systems of representation. One is a small number system, capable of delivering exact, discrete representations and calculations (Trick & Pylyshyn, 1993, 1994; Dehaene & Changeux, 1993; Starkey, 1992; Klein & Starkey, 1987; Honig & Matheson, 1995; Starkey, Spelke, & Gelman, 1990; Dehaene, 1997). It may be speculated that the preverbal, small number, exact system "module" (as Fodor (1975) calls it in his thesis, <u>The Modularity of Mind</u>), teaches us to track or localize numbers. For instance, it is well known that children, adults, and animals can track three items preverbally and discriminate them exactly from one or two, even if heterogeneously composed (Starkey, Spelke, &

Gelman 1983, 1990). It is also well understood that the "subitization" phenomenon describes how even infants (e.g., Pylyshyn, 1994; Pylyshyn et al., 1994) discriminate small numerosities exactly, which may be linked to "localizing and tracking objects in space" (Trick & Pylyshyn, 1993).

The second system is a large number system, approximating representations as vaguely somewhere 'out there', either as a blur somewhere between two apparently distantly spaced numbers, or as a blur somewhere beyond small numbers (Dehaene, 1992; Trick & Pylyshyn, 1994; see Dehaene, 1997 for comprehensive discussion). Well documented is the ability of adults, children, and animals to approximate large numerosities (Gelman & Meck, 1983; Bloom, 1994; Wynn & Bloom, 1992; Bloom & Wynn, 1997; Dehaene, 1997; see also Putnam, 1975).

The present studies offer persuasive evidence that human adults make use of an additional system which by yoking concepts and representations from both systems, enables them to verbalize large, exact numbers without limit and perform large, exact calculations. Certainly this sort of work has its theoretical roots in Chomsky's notion of a "generative grammar" (Chomsky, 1966) giving rise to the human conceptualization and verbalization of infinitely large numbers exactly (Bloom, 1994). Threading from Chomsky's theory of "Universal Grammar" (Chomsky, 1975) with which every student of cognitive science is well acquainted, is the notion that humans develop to become capable of 'cognizing' infinitely large numbers, the rules for which are innately known, the application of which awaits performance in language. In this paper, we lend additional support to Chomsky's foundations as we argue that language mediates the mapping that allows for human adults' unique facility for exact, large calculations and representations (Spelke & Tsivkin, in press). In so doing, we will trace structuralist[3] accounts of how this mapping might occur.

A Speculative Thesis

As Liz Spelke points out in the first articulation of her theory of conjoined modular systems, support may be obtained from two distinct fronts: studies with neurologically impaired subjects and studies such as this one involving bilingual subjects (Spelke & Tsivkin, in press). What the latter bring into focus are the mechanisms of mapping, that is in this case, how codes interrelate within a functional system to explain phenomena such as large exact calculations in human adults. What discussions of such studies almost always neglect, however, is an examination of their own basic assumptions. Why is it, for example, that we take a given code in one language to correspond neatly to a given code in another? Do symbols ever precisely stand for other symbols and nothing else? At stake in the present set of experiments is whether large, discrete numerosities are represented preverbally before they emerge verbally; yet this presupposes that such neat, unobstructed mappings are possible. In sketching a framework with which to understand conjoined modularities, therefore, we will attempt to understand representation itself.

For this, we turn to the lexicon and work of structural linguists, most notably Ferdinand de Saussure who is commonly identified with semiotics or the science of signs. By doing so, we make no claim that structuralism offers the only matrix in which to understand the behavioral phenomena we have observed. We assert only that it offers an interesting perspective at the very least, and one that inasmuch as it accounts for the phenomena, provides insight into their workings.

Saussure (1959) understood language as constituted by, of, and through signs that emerge from the relation of "signifiers" and "signifieds" (Greimas & Courtes, 1982) which are themselves signs.[4] Above all, the "sign" is arbitrary and differential, as Saussure himself put it:

> . . . in language there are only differences . . . Whether we take the signified or signifier, language has neither ideas nor sounds that existed before the linguistic system, but only conceptual and phonic differences that have issued from the system. (Saussure, 1959, p. 117-118, 120)

So, from this perspective, it is not that the number '3' symbolizes any quality of '3ness' or that it bears any ineluctable connection with the value or content of '3'. '3' is nothing more or less than an arbitrary symbol/sign for '3' (Eisele, 1985; Pimm, 1991).

Hence, the sign becomes meaningful (i.e., functional) via the play of differences between the network of signifiers and signifiers by which it has been constituted, and via its difference from other signs (Grosz, 1990; see also Peirce, 1966; Barthes, 1972; and Jakobson, 1975, cited in Sebeok, 1979). Understood as a sign, then, the number '3' means '3' only because it is different from '4' and/or '040272'. And if these signs emerge as a complex of signs as has been described thus far, then the '5' in '5 apples' cannot possibly be the same as the '5' in '5 songs'; nor can the '12' in '12 dogs' be equivalent to the '12' in the French '12 *chiens*'. All relations emerge together to differentiate and lend structural meaning/functionality to the sign.

Even at this point in our discussion of the sign, it should be apparent that the results of Experiment 3 are supported by this system. If a sign has meaning only in relation to the totality of other signs (i.e., signifiers and signifieds) in its domain, and is conditioned through the multiplicity of their possible interrelations, then the system of relations we call a 'story' or 'lesson' should produce signs contingent on those signs for their reproduction (i.e., by memory). Experiment 3 supports this conclusion. What should not be clear at this point is why only the numerical and not non-numerical facts that were embedded in these stories were recalled language-dependently.

The remaining question raised by Experiment 3, along with the results of Experiments 1 and 2, are not as easily supported at this juncture but will be so shortly with a brief elaboration of semiotic theory. Becker and Varelas (1993) trace their anatomy of a sign-system in the domain of large numerical representation. They point out that using a place-value system and doing large number calculations requires "semiotic reversibility" (p. 425), meaning the ability to "bring the complete value of the individual sign forms from the

background to the foreground as necessary [and vice-versa]." (Becker & Varelas, 1993, p. 425)

Please note that "complete value" does not at all denote anything like 'true value'; more accurately, it is something more than the symbol and less than what may be loosely called the pure meaning (to the extent that there is such a thing). So, for instance, the complete value of the number-sign "423" could be 4 hundreds, 2 tens, and 3 ones; 4 hundreds and 23 ones; 423 ones and so on. The complete value is actually all of these possibilities simultaneously, with these possibilities functioning as signs themselves in relation to other signs along a continuous chain. What Becker and Varelas (1993) are suggesting is that to make efficient use of the number-sign "423", we must place all of the possibilities constituting the complete value of the sign into the background of reference. Functioning strictly as symbol, the number-sign becomes fully constituted, fully articulatable, and most importantly, fully functional within, say, a calculation.

The concept of "semiotic reversibility" implies that the larger the numbers and the more exact the calculations to which those numbers are applied, the greater the semiotic demand placed on the calculator who must revolve these relations alternatively from the foreground to the background of her field of reference. Another way of thinking about this is that the larger and more exact the numerosity, the more critical the mediation of language: without a sign system (i.e., language) by which complete value is reversed into the background of reference, large numerosities cannot be expressed in any functional or meaningful way.

Multiplication makes these mechanisms all the more salient as "sign-sign relations" are foregrounded (Becker & Varelas, 1993, p. 429) Unlike addition in which the interplay of signs may be enriched with the more familiar and tractable sign-sign relations constituted by counting, subitization, and elemental numerosity discriminations (see Gallistel & Gelman, 1992; please see also Sophian, Wood, & Vong, 1995, and Wynn, 1995 for "relational" accounts of numerical representation without counting or counting-on

components), multiplication may be termed 'hyper-structural' as a semiotic performance.[5] After all, the multiplication tables we all learned as children contain nothing more than "arbitrary facts" as Dehaene (1997, p. 126) put it.

As a system of signs that makes obvious its arbitrariness, multiplication develops far more 'promiscuous' sign-sign relations (Campbell, 1987; Ashcraft, 1993): It is not uncommon to find adults confusing the products of sets of factors in the multiplication tables they labored to learn as children. Since the sign-sign relations in multiplication reference only themselves (i.e., they are 'hyper-structural'), they make no reference to a 'background' or other constitutive relations that might help them distinguish factor-set from factor-set.

Naturally a system of signs so invested in itself must show dependence on the language giving rise to those signs. This explains why Experiment 2 revealed language-specificity for both approximate and exact multiplication.

This semiotic framework lays down a powerful theoretical infrastructure for Liz Spelke's thesis that language mediates the conjunction of a small, exact calculation-capable number system and a large, approximation-capable number system. Ultimately, this framework speaks to computational, cognitive, and representational economy.

In the concurrent veins of artificial intelligence, computer systems, and cognition, Vera and Simon (1993) put forward what we may term a fundamental principle of advanced, efficient processing: Whether the processor be a human symbol processor as Pylyshyn would have it (Pylyshyn, 1984), or a mechanical one, it must be "functionally transparent" for it to manipulate large data inputs economically. The processor must, in this sense, operate at a "high functional level without the need for conscious awareness of symbolic representations at lower functional levels" (Vera & Simon, 1993, p. 14); it must function at the 'meta' or structural level, without drilling down into what Vera and Simon (1993) termed "details".

In the context of numerical representation and arithmetic, we hesitate to designate numerical value (i.e., the preverbal, mental representation of number) as "details" but this is precisely what they become if in manipulating large numbers we must pause to consider, for instance, place value as truly the value of the place. If in this sense, we privilege the signified over the signifier, that is, the content/value over the structure/sign, then the economy of our representational system is impeached. Only at the meta-level--at the level of the sign or the symbol itself--can large numbers be manipulated efficiently. This is the level mediated and enacted by language.

In The Modularity of Mind, Jerry Fodor (1983) makes a comparable argument about neural economy and language processing. It is worth quoting in full here:

> . . . input-processing for language provides no semantic analysis 'inside' lexical items. Or, to put it another way, the functionally defined level *output of the language processing module* respects such *structurally* defined notions as *item in the morphemic inventory of the language* . . . What we find . . . is that the fast, mandatory, [and "informationally encapsulated"] processes deliver representations of utterances which make perfectly good sense considered *as* representations of utterances; representations which specify, for example, morphemic constituency, syntactic structure, and logical form. (pp. 92-3, underline mine)

What Fodor (1983) has sketched, revised within a modular view of cognition, is what semioticians have always already known to be true of language: that if it is meaningful at all (i.e., if it is functional), it means structurally, through the play of differences and via its value as representation and representation only. So, in seeking an economical apparatus with which to manipulate large numbers exactly, human adults might unastonishingly make use of a system already configured for the job: one that is fast, efficient, and operates, as Fodor (1983) put it, "at the surface level" of the utterance by "ignor[ing] lots of the facts" (p. 70; we should recall here Vera and Simon's (1993) discussion of an economical AI system that would ignore the "details").[6]

Language-Specificity vs. Language-Dependence

The present studies reveal that bilingual subjects who learned new arithmetic facts requiring exact calculations, performed better in the language in which those facts were learned. These findings confirm anecdotal evidence suggesting that bilinguals who learn to manipulate large discrete numerosities in L1, prefer to continue to do so in L1 throughout their lives, even if they have spoken nothing but L2 for decades. However, the theory of conjoined modularities and the linguistic framework that explains it, does not account fully for our findings. While our theory supports language-dependence, our findings support both language-dependence *and* language-specificity.[7] The theory requires only that conjoined preverbal systems locate a system of signs in which large exact numerosities may be verbalized metalinguistically: it does not specify any particular system of signs. In this view, L2 (or, the language in which no training occurred) should do the job as well as L1 (or, the language of training). Of course, the results of our experiments do demonstrate language-specificity: subjects perform large exact calculations faster and more accurately in the language in which those calculations were learned within the experimental setting. In concluding this paper, we will examine more closely the semiotics of numerical representation in L1; we will determine why the representation of large exact numerosities *specifies* the L1 sign-system.

In drilling down into the dynamics of representation, we encounter numerous philosophical challenges (Pylyshyn, 1984). We have already seen that signs--the basic elements of language--do not mean any one thing but rather achieve meaningfulness or functionality inasmuch as they are different from other signs.[8]

Consider, for instance, the distinction drawn by Cheney and Seyfarth (1990) in their comprehensive exploration of the cognition of monkeys: The authors point out that the same sign--an "*association*" or habitual gathering of 2 monkeys--can be understood by the observing monkey (or human, for that matter) on one level as simply an "observable

fact" but on another level as a function of a complex social relationship implicating everything from the age, gender, and/or family hierarchies of the participants.

It may be argued that even were a monkey to articulate in English her understanding of the "association" as the mating, for example, of a high-ranking male and female, is the content of her mental representation then limited to this observable fact or is it composed of other fibers, equally resonant? Does this sort of 'fact', so to speak, ever emerge alone? Or, do sign, referent, and vocalization emerge together with their 'meaning' arising from the totality of meta-facts (e.g., the monkey's anxiety about what she's seeing; the intersection of the monkey's previous experience with what she's seeing; the monkey's premonitions of what she will do with the information she's acquired)?

As the above example illustrates, signs emerge "complexly conditioned" by the cues/signs that give rise to them (Linhart, 1989; see also Pylyshyn, 1984, and Campbell, 1987). In focusing on children's "linguistic cues", Wynn, Bloom (1992, 1997), and others (Johnson, 1997; Garson, 1994; Pylyshyn, 1984) have charted specific cues with which, in their words, "number word meaning" is mapped onto the codes by which they are represented. Citing "syntactic" and "semantic" cues, and delineating their properties (e.g., "Number words can only be used with count nouns, not with mass nouns"; and "Number words cannot appear with modifiers" as examples; Wynn & Bloom, 1992, pp. 516-7), Wynn and Bloom (1992), and then later, Bloom and Wynn (1997) effectively dismantle the one-to-one conception of code to content for numerical representation.

We may now pause here to return to the question posed earlier concerning the results of Experiment 3. If number-signs embedded in history or geography lessons, emerge conditioned by the cues embedded in those texts, then, we should not be surprised to find that they show language-specificity. Why number signs should be any different from other signs (non-numerical facts) also embedded and thereby conditioned in the texts, should now be clear: Number signs are distinctively cued. The network of signifying relations engendered by these cues and conditioning the number signs, remains uniquely

attached to these signs (i.e., it does not form "promiscuous" attachments; cf. earlier discussion of multiplication and 'promiscuity'). Hence we find language-specific recall of text-embedded numerical facts. Wynn and Bloom (1992), and others showed that the acquisition of numerical concepts is no more or less dependent on the cues with which they were learned, than on the concepts themselves. The concepts (of course, signs themselves) emerge together with these cues and associations, to become the numerical word/sign.

Hence, subjects who are trained in one language to perform large exact calculations and are then tested in another, perform better in the language of training. The language of training is the language in which she acquired not only the numerical word-signs but the system of signs/cues giving rise to those signs. The language in which new numerical facts are learned thereby becomes indispensable in the verbal configuration of those facts. As a result, the representation of large exact numerosity becomes more than language-dependent: it becomes language-specific.

Conclusion

Humans make use of a small number system module by which they discriminate and calculate small numerosities exactly. Humans make use of a large number system module by which they can generate infinite numerosities discretely and approximate calculations with large numerosities. Both of these modules operate language-independently or preverbally, and are fast and informationally encapsulated.

In order to calculate large numerosities exactly, humans make use of a third system that conjoins the mechanics of the small and large number system modules. This system performs language-dependently so that it may simulate the operational efficiency of the preverbal modules. Both the mechanics of these modules and their conjunction may be interrogated by exploring structuralist theories of language and thought.

The picture sketched out in this paper of what conjoined modularities might look like is indeed descriptive but may be viewed normatively as well. If the manipulation of

large exact numerosities efficiently and accurately is contingent upon a system of signs conducive to representational economy, then perhaps less emphasis should be placed on the rote memorization of multiplication tables and other tedious rote arithmetic. Instead educators might benefit from teaching their students to rely upon handheld calculators (this idea has also been proposed by Dehaene, 1997). Calculators do not fall prey to promiscuous associations and would keep students' numerical representational processing at the level of sign instead of between and among the signifiers that have constituted it.

Coextensive with this suggestion is that educators should teach algebra earlier. Like a calculator, algebra operates at the level of sign, and like multiplication, it makes obvious what is always already true of the representation and calculation of numerosity: that it is meaningful only insofar as it stands in relation to/difference from the other signifiers with which it is associated. Unlike multiplication, however, algebra is not challenged by the promiscuity of signifying relations since the relations that constitute its signs are articulated even in the more complex signs that it constructs.

For instance:

in **6 x 9 = 54**, what has happened to the signs 6 and 9 has evaporated in the product, 54;

but in **a x b = ab**, what has happened to the signs a and b is present, though transformed in ab.

The sign-sign relations in the latter remain intact. More interesting, these relations have been economized as a new, compact sign. In algebraic constructions, semiotic reversibility becomes unnecessary as foreground and background representations collapse onto each other to form an efficient, fast, and informationally encapsulated system of signs. An education that would emphasize algebra over more cumbersome manipulations of numerosities, might take mathematics education and its students one step higher in their evolving ability to cognize numerosity.

[1] This effect may be partially explained by the differing lengths of the verbal problems for the three sets of tasks (on average, 68.3, 61.7, and 86.2 characters for the approximate, novel base, and base-ten tasks, respectively).

[2] All the analyses presented below that involve numerical questions were initially compiled with the presumption that large-split and small-split numerical questions comprise two separate categories. However, since none of the analyses indicated any significant differences between large-split and small-split numerical questions, they were collapsed into one numerical category of questions which we then analyzed and presented below.

[3] Interestingly enough, the nexus between Chomsky's linguistics and traditionally structuralist linguistics is robust. For while only the latter is explicitly structuralist, both are inherently structural, and hence have much to contribute to each other. (Deely, 1990; Sebeok, 1994; Sturrock, 1986; Nuyts, 1993; Leiber, 1978; Caws 1988; Clarke, 1981)

[4] Saussure called the "signifier" that portion of the sign which is phonic or graphic, and "signified" the conceptual or functional component (Grosz, 1990; see also Holdcroft, 1991).

[5] Of course, this sort of neologism is moot: semiotics posits that the meaning of all linguistic relations and hence its inscription as language is entirely structural as the differential play of signs.

[6] In his treatise on "Language in Literature," Roman Jakobson (1987) comments in passing that children's acquisition of language is contingent on their acquisition of a "metalanguage", one that enables them to speak "about" a language--to use, in other words, more economically configured codes for large assemblies of signs. Our argument with respect to the representation of large numerosities is comparable: without language which is, to say, a "metalanguage" in Jakobson's terms, large numerical representation is not possible. (For a discussion of the cultural universality of metalinguistic acquisition, please see Carol Fleisher Feldman's discussion of "Oral Metalanguage" [Feldman, 1991]).

[7] The terms 'language-specificity' and 'language-dependence' are used interchangeably up until this point throughout the paper. It is only in this section that we tease the concepts apart a bit. This should not at all, however, diminish the interchangeability of the terms thus far.

[8] Please see an interesting discussion of the transformative nature of language by Linhart (1989).

Table 1.
An example of the display presented to subjects in Experiment 1.

How much is
Forty-Seven
plus
Fifty-Four?

ninety-one	one hundred one

Table 2.
Mean reaction time (RT) and error rate (ER) for the subjects' performance during the training sessions in Experiment 1.

	Addition in the Familiar Base Task				Addition in Unfamiliar Bases Task			
	Trained in Russian		Trained in English		Trained in Russian		Trained in English	
	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2
RT	4302	3106	4462	3255	3826	2825	3879	2756
ER	3%	1%	3%	1%	4%	1%	4%	1%

	Estimation Tasks			
	Trained in Russian		Trained in English	
	Day 1	Day 2	Day 1	Day 2
RT	3582	2624	3504	2545
ER	4%	1%	4%	1%

Table 3a.
Mean reaction time (RT) and error rate (ER) for the subjects' performance with trained ("old") items during testing in Experiment 1.

	Addition in Familiar Base Task				Addition in Unfamiliar Bases Task			
	Trained in Russian		Trained in English		Trained in Russian		Trained in English	
	Tested in Russian	Tested in English	Tested in English	Tested in Russian	Tested in Russian	Tested in English	Tested in English	Tested in Russian
RT	2598	2835	1986	2655	2027	2592	2909	3324
ER	0%	2%	1%	1%	1%	1%	0%	1%

	Estimation Task			
	Trained in Russian		Trained in English	
	Tested in Russian	Tested in English	Tested in English	Tested in Russian
RT	2629	2741	2661	2598
ER	0%	1%	0%	1%

Table 3b.
Mean response time (RT) and error rate (ER) for the subjects' performance on untrained ("new") items during testing in Experiment 1.

	Addition in Familiar Base Task				Addition in Unfamiliar Bases Task			
	Trained in Russian		Trained in English		Trained in Russian		Trained in English	
	Tested in Russian	Tested in English	Tested in English	Tested in Russian	Tested in Russian	Tested in English	Tested in English	Tested in Russian
RT	2973	3827	2875	3616	3215	3564	3256	3123
ER	1%	1%	1%	1%	1%	1%	0%	2%

	Estimation Task			
	Trained in Russian		Trained in English	
	Tested in Russian	Tested in English	Tested in English	Tested in Russian
RT	2671	2698	2691	2708
ER	1%	1%	0%	0%

Table 4.
An example of the displays presented to the subjects in Experiment 2.

How much is
Fifty-Eight
plus
Sixty-Three?

one hundred twenty-one	one hundred thirty-one

Estimate
How much is
Fifty-Eight
plus
Sixty-Three?

one hundred thirty	one hundred

Table 5.
Mean reaction time (RT) and error rate (ER) for the subjects' performance with addition tasks during the training sessions in Experiment 2.

	Approximate Addition Task				Exact Addition Task			
	Trained in Russian		Trained in English		Trained in Russian		Trained in English	
	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2
RT	4461	2277	4385	2458	4227	2790	4343	2835
ER	3%	0%	3%	1%	4%	1%	3%	1%

Table 6a.
Mean reaction time (RT) and error rate (ER) for subjects' performance with trained ("old") addition items during testing in Experiment 2.

	Approximate Addition Task				Exact Addition Task			
	Trained in Russian		Trained in English		Trained in Russian		Trained in English	
	Tested in Russian	Tested in English	Tested in English	Tested in Russian	Tested in Russian	Tested in English	Tested in English	Tested in Russian
RT	3221	3147	3264	3285	3250	4002	2677	4023
ER	0%	1%	1%	1%	1%	1%	1%	1%

Table 6b.
Mean reaction time (RT) and error rate (ER) for subjects' performance with untrained ("new") addition items during testing in Experiment 2.

	Approximate Addition Task				Exact Addition Task			
	Trained in Russian		Trained in English		Trained in Russian		Trained in English	
	Tested in Russian	Tested in English	Tested in English	Tested in Russian	Tested in Russian	Tested in English	Tested in English	Tested in Russian
RT	3226	2904	3361	3356	3921	4391	3600	4473
ER	1%	1%	1%	1%	1%	1%	1%	1%

Table 7.
Mean reaction time (RT) and error rate (ER) for subjects' performance with trained ("old") addition items in both answer formats during testing in Experiment 2.

TRAINED ON APPROXIMATE ADDITION PROBLEMS								
	Tested with Approx. Answers				Tested with Exact Answers			
	Trained in Russian		Trained in English		Trained in Russian		Trained in English	
	Tested in Russian	Tested in English	Tested in English	Tested in Russian	Tested in Russian	Tested in English	Tested in English	Tested in Russian
RT	3221	3147	3264	3285	3474	3412	3713	3799
ER	0%	1%	1%	1%	1%	1%	1%	2%

TRAINED ON EXACT ADDITION PROBLEMS								
	Tested with Exact Answers				Tested with Approx. Answers			
	Trained in Russian		Trained in English		Trained in Russian		Trained in English	
	Tested in Russian	Tested in English	Tested in English	Tested in Russian	Tested in Russian	Tested in English	Tested in English	Tested in Russian
RT	3250	4002	2677	4023	3221	4031	2723	3646
ER	1%	1%	1%	1%	1%	2%	1%	1%

Table 8.
Mean reaction time (RT) and error rate (ER) for the subjects' performance with multiplication tasks during the training sessions in Experiment 2.

	Approximate Multiplication Task				Exact Multiplication Task			
	Trained in Russian		Trained in English		Trained in Russian		Trained in English	
	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2
RT	3950	2914	4493	2895	5823	3495	5173	2765
ER	3%	1%	4%	1%	3%	1%	3%	1%

Table 9a.
Mean reaction time (RT) and error rate (ER) for subjects' performance with trained ("old") multiplication items during testing in Experiment 2.

	Approximate Multiplication Task				Exact Multiplication Task			
	Trained in Russian		Trained in English		Trained in Russian		Trained in English	
	Tested in Russian	Tested in English	Tested in English	Tested in Russian	Tested in Russian	Tested in English	Tested in English	Tested in Russian
RT	2614	2836	2874	3333	3900	5337	4971	5497
ER	2%	2%	2%	2%	0%	1%	1%	1%

Table 9b.
Mean reaction time (RT) and error rate (ER) for subjects' performance with untrained ("new") multiplication items during testing in Experiment 2.

	Approximate Multiplication Task				Exact Multiplication Task			
	Trained in Russian		Trained in English		Trained in Russian		Trained in English	
	Tested in Russian	Tested in English	Tested in English	Tested in Russian	Tested in Russian	Tested in English	Tested in English	Tested in Russian
RT	3469	3454	3288	3836	5568	5487	5301	6425
ER	2%	2%	2%	2%	1%	1%	1%	1%

Table 10.
Mean reaction time (RT) and error rate (ER) for the subjects' performance during the training sessions in Experiment 3.

Story A ("Kapnopa")						
	Trained in Russian			Trained in English		
	Day 1	Day 2	Day 3	Day 1	Day 2	Day 3
RT	4277	3700	2688	4544	3521	2720
ER	22%	10%	2%	15%	12%	0%

Story B ("Mary")						
	Trained in Russian			Trained in English		
	Day 1	Day 2	Day 3	Day 1	Day 2	Day 3
RT	4763	3464	2962	4503	3515	2640
ER	25%	10%	0%	15%	7%	2%

Table 11a.
Mean reaction time (RT) and error rate (ER) for subjects' performance with trained ("old") numerical and non-numerical facts during testing in Experiment 3.

	Large Number Facts				General Non-Numerical Facts			
	Trained in Russian		Trained in English		Trained in Russian		Trained in English	
	Tested in Russian	Tested in English	Tested in English	Tested in Russian	Tested in Russian	Tested in English	Tested in English	Tested in Russian
RT	2285	3143	2730	3036	2295	2324	2462	2380
ER	3%	3%	0%	3%	1%	3%	0%	1%

Table 11b.
Mean reaction time (RT) and error rate (ER) for subjects' performance with untrained ("new") numerical and non-numerical facts during testing in Experiment 3.

	Large Number Facts				General Non-Numerical Facts			
	Trained in Russian		Trained in English		Trained in Russian		Trained in English	
	Tested in Russian	Tested in English	Tested in English	Tested in Russian	Tested in Russian	Tested in English	Tested in English	Tested in Russian
RT	2352	3096	2758	3047	2220	2395	2484	2429
ER	3%	3%	3%	3%	1%	1%	0%	0%

Table 12a.
Mean reaction time (RT) and error rate (ER) for subjects' performance with trained ("old") small number facts during testing in Experiment 3.

	Small Fractions				Ordinal Number Facts			
	Trained in Russian		Trained in English		Trained in Russian		Trained in English	
	Tested in Russian	Tested in English	Tested in English	Tested in Russian	Tested in Russian	Tested in English	Tested in English	Tested in Russian
RT	2224	2898	2857	2666	2237	2909	2640	2510
ER	0%	6%	3%	0%	0%	0%	3%	3%

Table 12b.
Mean reaction time (RT) and error rate (ER) for subjects' performance with untrained ("new") small number facts during testing in Experiment 3.

	Small Fractions				Ordinal Number Facts			
	Trained in Russian		Trained in English		Trained in Russian		Trained in English	
	Tested in Russian	Tested in English	Tested in English	Tested in Russian	Tested in Russian	Tested in English	Tested in English	Tested in Russian
RT	2307	2965	2834	2639	2283	2895	2632	2594
ER	3%	3%	0%	3%	0%	3%	0%	6%

Table 13a.
Mean reaction time (RT) and error rate (ER) for subjects' performance with trained ("old") spatial and temporal facts during testing in Experiment 3.

	Egocentric Facts				Geocentric Facts			
	Trained in Russian		Trained in English		Trained in Russian		Trained in English	
	Tested in Russian	Tested in English	Tested in English	Tested in Russian	Tested in Russian	Tested in English	Tested in English	Tested in Russian
RT	2824	3264	3146	3033	2587	3454	2941	3117
ER	3%	6%	3%	6%	6%	6%	12%	6%

	Seasons				Time of Day			
	Trained in Russian		Trained in English		Trained in Russian		Trained in English	
	Tested in Russian	Tested in English	Tested in English	Tested in Russian	Tested in Russian	Tested in English	Tested in English	Tested in Russian
RT	2790	3055	2723	2741	2598	3209	3317	2722
ER	12%	12%	6%	6%	6%	6%	12%	12%

Table 13b.
Mean reaction time (RT) and error rate (ER) for subjects' performance with untrained ("new") spatial and temporal facts during testing in Experiment 3.

	Egocentric Facts				Geocentric Facts			
	Trained in Russian		Trained in English		Trained in Russian		Trained in English	
	Tested in Russian	Tested in English	Tested in English	Tested in Russian	Tested in Russian	Tested in English	Tested in English	Tested in Russian
RT	2847	3235	3070	3043	2692	3290	3113	2975
ER	6%	3%	3%	6%	0%	6%	6%	12%

	Seasons				Time of Day			
	Trained in Russian		Trained in English		Trained in Russian		Trained in English	
	Tested in Russian	Tested in English	Tested in English	Tested in Russian	Tested in Russian	Tested in English	Tested in English	Tested in Russian
RT	3116	3247	2871	3061	2647	3309	3361	2840
ER	12%	6%	12%	6%	6%	12%	12%	6%

References

Ashcraft, M. H. (1993). Cognitive arithmetic: A review of data and theory. In S. Dehaene (Ed.), Numerical Cognition (pp. 75-106). Cambridge, MA: Blackwell.

Barthes, R. (1972). Mythologies (A. Lavers, Trans.). New York: The Noonday Press. (Original work published 1957).

Becker, J., & Varelas, M. (1993). Semiotic aspects of cognitive development: Illustrations from early mathematical cognition. Psychological Review, 100 (3), 420-431.

Bloom, P. (1994). Generativity within language and other cognitive domains. Cognition, 51, 177-189.

Bloom, P., & Wynn, K. (1997). Linguistic cues in the acquisition of number words. Journal of Child Language, 24, 511-533.

Boysen, S. T., & Berntson, G. G. (1989). Quantity-based interference and symbolic representations in chimpanzees (*Pan trogolodytes*). Journal of Experimental Psychology: Animal Behavior Processes, 22, 76-86.

Boysen, S. T., & Capaldi, E. J. (Eds.) (1993). The development of numerical competence: Animal and human models. Hillsdale, NJ: Erlbaum.

Campbell, J. I. D. (1987). The role of associative interference in learning and retrieving arithmetic facts. In J. A. Sloboda & D. Rogers (Eds.), Cognitive Processes in Mathematics (pp. 107-122). Oxford: Oxford University Press.

Carey, S., & Spelke, E. S. (1994). Domain-specific knowledge and conceptual change. In L. A. Hirschfeld & S. A. Gelman, Mapping the mind. Cambridge: Cambridge University Press.

Caws, P. (1988). Structuralism: The art of the intelligible. Atlantic Highlands, NJ: Humanities Press International.

Cheney, D. L., & Seyfarth, R. M . (1990). How monkeys see the world. Chicago: University of Chicago Press.

Chomsky, N. (1966). Topics in the Theory of Generative Grammar. The Hague: Mouton.

Chomsky, N. (1975). Reflections on language. New York: Pantheon Books.

Chomsky, N. (1986). Knowledge of language: Its nature, origin, and use. New York: Praeger.

Clarke, S. (1981). The foundations of structuralism. Sussex: The Harvester Press.

Davis, H. (1993). Numerical competence in animals: Life beyond Clever Hans. In S. T. Boysen & E. J. Capaldi (Eds.), The development of numerical competence: Animal and human models (pp. 109-125). Hillsdale, NJ: Lawrence Erlbaum Associates.

Davis, H., & Perusse, R. (1988). Numerical competence in animals: Definitional issues, current evidence, and a new research agenda. Behavioral and Brain Sciences, 11, 561-615.

Deely, J. (1990). Basics of semiotics. Bloomington: Indiana Universtiy Press.

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

Dehaene, S. (1997). The number sense: How the mind creates mathematics. New York: Oxford University Press.

Dehaene, S., & Changeux, J. P. (1993). Development of elementary numerical abilities: A neuronal model. Journal of Cognitive Neuroscience, 5 (4), 390-407.

Dehaene, S., & Cohen, L. (1991). Two mental calculation systems: A case study of severe acalculia with preserved approximation. Neuropsychologia, 29 (11), 1045-1074.

Eisele, C. (1985). Historical perspectives on Peirce's logic of science. Berlin: Mouton Publishers.

Ellis, N. (1980). Linguistic relativity revisited: The bilingual word-length effect in working memory during counting, remembering numbers, and mental calculation. In R. J. Harris (Ed.), Cognitive processing in bilinguals. Amsterdam: Elsevier.

Feldman, C. F. (1991). Oral metalanguage. In D. R. Olson & N. Torrance (Eds.), Literacy and orality (pp. 47-65). Cambridge: Cambridge University Press.

Fodor, J. A. (1975). The language of thought. New York: Crowell.

Fodor, J. A. (1983). The modularity of mind. Cambridge, MA: Bradford Books/MIT Press.

Frenck-Mestre, C., & Vaid, J. (1993). Activation of number facts in bilinguals. Memory and Cognition, 21 (6), 809-818.

Gallistel, C. R. (1990). The organization of learning. Cambridge, MA: Bradford Books/MIT Press.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. Cognition, 44, 43-74.

Garson, J. W. (1994). No representtions without rules: The prospects for a compromise between paradigms in cognitive science. Mind and Language, 9 (1), 25-37.

Gelman, R., & Gallistel, C. R. (1978). The child's understanding of number. Cambridge, MA: Harvard Universtiy Press.

Gelman, R., & Meck, E. (1983). Preschoolers' counting: Principles before skill. Cognition, 13, 343-359.

Greimas, A. J., & Courtes, J. (Eds.) (1982). Semiotics and language: An analytical dictionary (L. Crist, D. Patte, J. Lee, McMahon, E., G. Phillips, M. Rengstorf, Trans.). Bloomington: Indiana University Press. (Original work published in 1979)

Gonzalez, E.G., & Kolers, P. A. (1987). Mental manipulation of arithmetic symbols. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8 (4), 308-319.

Grosz, E. (1990). Jacques Lacan: A feminist introduction. New York: Routledge.

Hauser, M, & Carey, S. (in press). Building a cognitive creature from a set of primitives: Evolutionary and developmental insights. In C. Allen & D. Cummins (Eds.), The Evolution of Mind. Oxford: Oxford University Press.

Hauser, M., McNeilage, P., & Ware, M. (1996). Numerical representations in primates. Proceedings of the National Academy of Sciences, 93, 1514-1517.

Hermer, L., & Spelke, E. S. (1996). Modularity and development: The case of spatial reorientation. Cognition, 61 (3), 195-232.

Holdcroft, D. (1991). Saussure: Signs, system, and arbitrariness. Cambridge: Cambridge University Press.

Honig, W. K., & Matheson, W. R. (1995). Discrimination of relative numerosity and stimulus mixture by pigeons with comparable tasks. Jounal of Experimental Psychology: Aminal Behavior Processes, 21 (4), 348-363.

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

Jacobson, R. (1975). Coup d'oeil sur le developpement de la semiotique. Studies in Semiotics 3. The Hague: Mouton.

Jacobson, R. (1987). Language in literature. Cambridge, MA: Harvard University Press.

Johnson, M. H. (1997). The neural basis of cognitive development. In W. Damon, D. Kuhn & R. Siegler (Eds.), Handbook of Child Psychology: Vol. 2. Cognition, Perception, and Language (pp. 1-49). New York: John Wiley & Sons, Inc.

Klein, A., & Starkey, P. (1987). The orinigs and development of numerical cognition: A comparative analysis. In J. A. Sloboda & D. Rogers (Eds.), Cognitive Processes in Mathematics (pp. 1-25). Oxford: Oxford University Press.

Kolers, P. A. (1968). Bilingualism and information processing. Scientific American, 218 (3), 78-86.

Leiber, J. (1978). Structuralism. Boston, MA: Twayne Publishers.

Levinson, S. C. (1996). Relativity in spatial conception and description. In J. J. Gumperz & S. C. Levinson (Eds.), Rethinking linguistic relativity: Studies in the social and cultural foundations of language, No. 17 (pp. 177-202). Cambridge: Cambridge University Press.

Linhart, J. (1989). New categorial system and cognitive psychology. Studia Psychologica, 31 (3), 171-196.

McClain, L, & Huang, J. Y. S. (1982). Speed of simple arithmetic in bilinguals. Memory and Cognition, 10 (6), 591-596.

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

Marsh, L. G., & Maki , R. H. 1976). Efficiency of arithmetic operations in bilinguals as a function of language. Memory and Cognition, 4 (4), 459-464.

Matsuzawa, T. (1985). Use of numbers by a chimpanzee. Nature, 315, 57-59.

Nuyts, J. (1993). Cognitive linguistics. Journal of Pragmatics, 20, 269-290.

Peirce, C. S. (1935-1966). Collected papers of Charles Sanders Peirce (ed. by C. Hartshorne, P. Weiss & A. W. Burks). Cambridge, MA: Harvard University Press.

Pepperberg, I. M. (1987). Evidence for conceptual quantitative abilities in the African grey parrot: Labeling of cardinal sets. Ethology, 75, 37-61.

Pimm, D. (1991). Communicating mathematically. In K. Durkin & B. Shire, Language in mathematical education (pp. 17-23). Philadelphia: Open University Press.

Putnam, H. (1975). Mind, language and reality. (Vols. 1-2). Cambridge: Cambridge University Press.

Pylyshyn, Z. W. (1984). Computation and cognition. Cambridge, MA: Bradford Books/MIT Press.

Pylyshyn, Z. W. (1994). Some primitive mechanisms of spatial attention. Cognition, 50, 363-384.

Pylyshyn, Z. W., Burkell, J., Fisher, B., Sears, C., Schmidt, W., & Trick, L. (1994). Multiple parallel access in visual attention. Canadian Journal of Experimental Psychology, 48 (2), 260-283.

Rilling, M. (1993). Invisible counting animals: A history of contributions from comparative psychology, ethology, and learning theory. In S. T. Boysen & E. J. Capaldi (Eds.), The development of numerical competence: Animal and human models (pp. 3-37). Hillsdale, NJ: Lawrence Erlbaum Associates.

Saussure, F. de (1959). Course in general linguistics (W. Baskin, Trans.). New York: Philosophical Library.

Sebeok, T. A. (1979). The sign and its masters. Austin: Universtiy of Texas Press.

Sebeok, T. A. (1994). Signs: An introduction to semiotics. Toronto: University of Toronto Press.

Sophian, C., Wood, A. M., & Vong, K. I. (1995). Making numbers count: The early development of numerical inferences. Developmental Psychology, 31 (2), 263-273.

Spelke, E. S., & Tsivkin, S. (in press). Initial knowledge and conceptual change: Space and number. To appear in Bowerman, M. & Levinson, S. C. (Eds.), Language acquisition and conceptual development. Cambridge: Cambridge University Press.

Starkey, P. (1992). The early development of numerical reasoning. Cognition, 43, 93-126.

Starkey, P., Spelke, E. S., & Gelman, R. (1983). Detection of intermodal numerical correspondences by human infants. Science, 222, 179-181.

Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numercial abstraction by human infants. Cognition, 36, 97-127.

Sturrock, J. (1986). Structuralism. London: Paladin Grafton Books.

Thomas, R. K., & Lorden, R. B. (1993). Numerical competence in animals: A conservative view. In S. T. Boysen & E. J. Capaldi (Eds.), The development of numerical competence: Animal and human models (pp. 127-147). Hillsdale, NJ: Lawrence Erlbaum Associates.

Trick, L., M., & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive procesing. Journal of Experimental Psychology: Human Perception and Performance, 19, 331-351.

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 (1), 80-102.

Uller, C., Carey, S., & Hauser, M. (1996, April). Numerical representations in cotton-top tamarins (*Saguinus oedipus oedipus*). Poster presentation at the International Conference on Infant Studies, Providence, RI.

Vera, A. H., & Simon, H. A. (1993). Situated action: A symbolic interpretation. Cognitive Science, 17, 7-48.

Warrington, E. K. (1982). The fractionation of arithmetical skills: A single case study. Quarterly Journal of Experimental Psychology, 34A, 31-51.

Wynn, K. (1990). Children's understanding of counting. Cognition, 36 (2), 155-193.

Wynn, K. (1995). Infants possess a system of numberical knowledge. Current Directions in Psychological Science, 4 (6), 172-177.

Wynn, K., & Bloom, P. (1992). The origins of psychological axioms of arithmetic and geometry. Mind and Language, 7 (4), 409-416.