ELSEVIER

COGNITION

Cognition 78 (2001) 45–88

www.elsevier.com/locate/cognit

Language and number: a bilingual training study

Elizabeth S. Spelke*, Sanna Tsivkin

Department of Brain and Cognitive Sciences, MIT, NE20-456, 3 Cambridge Center, Cambridge, MA 02139, USA

Received 15 June 1999; received in revised form 2 May 2000; accepted 16 August 2000

Abstract

Three experiments investigated the role of a specific language in human representations of number. Russian–English bilingual college students were taught new numerical operations (Experiment 1), new arithmetic equations (Experiments 1 and 2), or new geographical or historical facts involving numerical or non-numerical information (Experiment 3). After learning a set of items in each of their two languages, subjects were tested for knowledge of those items, and new items, in both languages. In all the studies, subjects retrieved information about exact numbers more effectively in the language of training, and they solved trained problems more effectively than untrained problems. In contrast, subjects retrieved information about approximate numbers and non-numerical facts with equal efficiency in their two languages, and their training on approximate number facts generalized to new facts of the same type. These findings suggest that a specific, natural language contributes to the representation of large, exact numbers but not to the approximate number representations that humans share with other mammals. Language appears to play a role in learning about exact numbers in a variety of contexts, a finding with implications for practice in bilingual education. The findings prompt more general speculations about the role of language in the development of specifically human cognitive abilities. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Language; Number; Bilingual training study

1. Introduction

Research over the last decades has provided evidence for representations of number in a variety of non-human animals (for reviews see Boysen & Capaldi, 1993; Davis & Pérusse, 1988; Dehaene, 1997; Gallistel, 1990). For example,

* Corresponding author. Fax: +1-617-258-8654.
E-mail address: spelke@mit.edu (E.S. Spelke).

0010-0277/01/$ - see front matter
PII: S0010-0277(00)00108-6

untrained monkeys represent the exact number of objects in a scene, provided the number is small, and take account of the effects of additions and subtractions of single objects (Hauser, Carey, & Hauser, 2000; Hauser, MacNeilage, & Ware, 1996). Trained and untrained birds, fish, and mammals represent the approximate numerosity of larger sets of items (for discussion see Gallistel, 1990). After extensive training, several chimpanzees and one parrot learned symbols for exact numbers of objects in sets as large as 10 (Boysen & Capaldi, 1993; Matsuzawa, 1985; Pepperberg, 1987), two monkeys learned to order the numbers 1–4 and gave ordinal judgements for numbers up to 8 (Brannon & Terrace, 1998), and one chimpanzee learned to use number symbols to enumerate objects under conditions suggesting a process of addition (Boysen & Berntson, 1989). In all these cases, the performance of non-human animals either equaled or exceeded the performance of human infants tested with no training (for discussion see Gallistel & Gelman, 1992; Hauser & Carey, 1998).

Although these abilities suggest considerable continuity in number representations over human evolution, there are striking discontinuities as well. Between the ages of 2 and 4, human children learn verbal counting (see Gelman & Gallistel, 1978; Wynn, 1990). 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 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 (for review see 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 investigators 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 explore 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, 1980). A third proposal is our focus here. Humans may have the same initial number capacities as other animals but may develop new number representations through the use of a specific language. The language faculty, operating in conjunction with the cognitive systems that humans share with other animals, underlies distinctively human knowledge of number.

Connections between knowledge of language and knowledge of number have been suggested on theoretical and empirical grounds. Chomsky (1986) noted that both the sentences of a language and the numbers in a counting sequence have the property of discrete infinity, and he suggested that the same recursive device underlies both (cf. Bloom, 1994; Hurford, 1987). Neuropsychologists have found that disorders in number representation frequently are accompanied by disorders in language

(Dehaene & Cohen, 1991; McCloskey, 1992; Warrington, 1982). 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 & Hennelly, 1980; Gathercole & Baddeley, 1993).

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 arithmetic. In some anecdotal cases, this tendency is both extreme and puzzling. For example, a person who learns elementary arithmetic in one language may move to a different language community, become dominant in the new language to the point of speaking and dreaming in that language and losing facility in the original language, and yet resort to the first language when adding up a bill or counting change (see Dehaene, 1997). In experiments, bilinguals have been found to solve arithmetic problems with greater speed and accuracy when the problems are presented in their first language (French-Mestre & Vaid, 1993; Gonzalez & Kolers, 1987; Kolers, 1968; Marsh & Maki, 1976; 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 calculated anew. The longer response times and lower accuracy at retrieving arithmetic facts in a second language therefore would stem either from a translation process or from less well-established 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 a spoken problem into a representation in the system in which the answer is computed, and then transform the result of the computation back into the spoken language for production. These decoding and encoding processes might proceed automatically, even when no spoken response is required, producing the language-specific effects described above (Holender & Peereman, 1987; McCloskey, 1992).

In the present experiments, we attempted to distinguish these possibilities by investigating whether different kinds of number facts are represented in a language-dependent or language-independent manner. We followed the tradition of investigating language and number through studies of bilingual learners, with two innovations. First, we conducted training studies in which bilingual 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 perform arithmetic or for both languages. It also allowed us to distinguish language-dependent number representations from language-specific encoding and decoding processes, because we could present subjects with different facts involving the same numbers in their two languages, giving equal training across languages to encoding and decoding of the number words.

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, exact number representations is language-dependent, whereas learning of new facts drawing on the representations humans share with other animals is not. By this hypothesis, large, exact number facts learned in one language should not be immediately accessible to queries in the other language, but facts about large, approximate numbers should be equally accessible to subjects regardless of the language in which they are queried.

We began with a study assessing bilingual learning of the exact results of large number additions in base 10, the exact results of additions in novel bases, and the approximate results of logarithmic and cube root functions (Experiment 1). Next we compared bilingual learning of new facts concerning 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 facts and non-numerical facts in fictitious history and geography lessons. After presenting these experiments, we propose an account for the observed patterns of language-dependence and language-independence in number representations and attempt to characterize, more generally, the role of language in creating representations that are unique to humans. We also sketch some possible implications of our research for contemporary debates about bilingual education.

2. Experiment 1

Russian–English bilingual college students were taught four sets of facts involving relations among large, exact numbers. Two sets of facts involved the familiar operation of addition in base 10: adding either 54 or 63 to each of a set of two-digit numbers. Two further sets of facts involved the less familiar operations of addition in base 6 and base 8. The students also were taught two sets of facts involving large, approximate numbers: the approximate cube roots and the approximate base 2 logs of each of a set of large numbers. Each student learned one set of exact large number addition facts, exact novel base addition facts, and approximate number facts in each language. Half the facts in each set involved specific numbers that appeared only in one language over the course of the experiment, and half the facts involved numbers that appeared in both languages. For the latter facts, exposure to the number words and practice at any decoding and encoding processes involving those words were equated across the two languages.

After 2 days of training (in which subjects' performance improved for all the problem sets), subjects were tested on all the facts 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 language-of-learning advantage for arithmetic calculation stems from processes that translate from the language of input to a language-independent system of representation, then subjects should perform better when tested in the language in which they learned and habitually perform arithmetic (Russian). Moreover, subjects should perform better in the language of training than in the untrained language for all facts involving monolingually-trained numbers. Repeated exposure to a number in one language should enhance the translation process for that number and language. In contrast, if the language-of-learning advantage for arithmetic calculations stems from the use of a language-specific system of representation, then performance in the language of training should exceed performance in the untrained language for all facts that require this system, whether the facts involved monolingually- or bilingually-trained numbers and whether the facts were learned in Russian or in English. We predicted that facts involving exact large numbers would show this language specificity, whereas facts involving approximate large numbers would not.

A secondary purpose of the experiment was to investigate whether subjects truly develop exact and approximate number representations by investigating whether training on specific items within a class of facts generalized to other facts within the class. 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 exact large addition problems in which a new number within the range of the training numbers was incremented by 54). Performance on these untrained problems was compared to performance on the trained problems, both in the language of training and in the untrained language. If subjects use approximate number representations to represent the new set of log and cube root facts, then their distinct approximate number representations should overlap (see Dehaene, 1997; Gallistel & Gelman, 1992), and learning should generalize to corresponding facts involving other numbers within the same range. In contrast, if subjects use exact number representations to represent the new sets of addition facts, then their representations of different number facts should not overlap and no such generalization should occur.

2.1. Method

2.1.1. Subjects

The 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). They were undergraduate or graduate students living in Ithaca, NY, and were solicited through the ethnic clubs on Cornell's campus. All subjects were native speakers of Russian, spoke no English before adolescence, spent at least 3 years in the US, and were comfortable conversing and reading in both Russian and English. The mean age at which the subjects started learning English was 15.4 years (range 13–19 years), and the mean time since coming to the US was 3.8 years (range 3–5 years). Subjects were required to pass comprehension tests in Russian and English before they were allowed to participate (see below). One additional subject was dropped from the study when the screening test revealed poor English comprehension.

2.1.2. Materials

The experiment was conducted on a Macintosh SE computer, with all the tasks appearing on the computer's 12 inch monitor. All stimuli appeared within a 512×600

pixel viewing area on the monitor. 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 English, all questions appeared in Geneva font size 48, and answers appeared in Geneva font size 28. For the stimuli in Russian, the Russian Central font was used in the same point sizes as for the English stimuli.

Stimuli consisted of six categories of arithmetic problems, described below, with one problem and two potential answers presented on the screen on 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 examples). 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 ms. Before the start of each set of training and test items, subjects were given three filler items to accustom or re-accustom them to the problems and procedure.

The six categories of problems were as follows (see Appendix A for examples):

(1) *Exact addition with addend 54*: subjects had 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 1, with the differences balanced in the underestimation and overestimation directions.[1]

(2) *Exact addition with addend 63*: subjects added 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 1.

(3) *Exact addition in base 6*: subjects learned the sums of 12 base 6 additions in which the first addend ranged from 1 to 3-3, the second addend ranged from 2 to 5-1, and the sum ranged from 3 to 1-2-1. The response alternatives were the correct answer and a distractor in which one of the digits was raised or lowered by 1 from the correct answer.

(4) *Exact addition in base 8*: subjects learned the sums of 12 base 8 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 to 1-1-5. The response alternatives again were the correct answer and a distractor in which one of the digits was raised or lowered by 1 from the correct answer.

(5) *Approximation of cube roots*: subjects learned to estimate the cube roots of 12 numbers ranging from 9 to 5830 by choosing the closer of two whole-number responses. When the correct answer was 4 or less, the distractor differed from it by 1 in either direction. When the correct answer was larger, the distractor differed from it by 2 in either direction.

[1] We chose these foils to ensure that subjects would perform the full two-digit addition; foils with an incorrect digit in the units place could have been detected after only the first step in the addition. Because instructions emphasized exact accuracy and because the two candidate answers differed by less than 10%, it is unlikely that any subjects attempted to choose the correct answer by estimation.

Table 1
Example problems for Experiment 1

Exact large addition	
What is the sum of fifty-four and forty-eight?	
One hundred two	Ninety-two
Exact novel base addition	
In base six, what is the sum of two and five?	
One-one	Two-one
Approximation	
Estimate the approximate cube root of twenty-nine.	
Four	Three

(6) *Approximation of logs base 2*: subjects learned to estimate the base 2 logarithms of 12 numbers ranging from 9 to 8250 by choosing the closer of two whole-number responses. When the correct answer was 4 or less, the distractor was 1 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 comprehension test composed of a variety of items with 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 both languages were within 500 ms of each other were permitted to continue with the experiment. One subject was dropped from the study when the comprehension test detected his marked superiority in Russian (mean RT = 2815 ms) over English (mean RT = 3550 ms).

2.1.3. 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 exact addition of 54, exact addition in base 6, and approximate cube root problems occurring during one session and exact addition of 63, exact addition in base 8, and approximate 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 six 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 two repetitions of each of six trained items, and two repetitions of each of six 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 could occur twice in succession. Each set of training and test problems was preceded by three filler items.

2.1.4. Procedure

All training sessions in each language were preceded by greetings and casual conversation in that language. Because all subjects lived and worked in an English-speaking environment, a further effort was made to re-accustom 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 and conversed informally with the experimenter about it.

For all training sessions each problem set began with instructions specific to that problem set and with example 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 the 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 ms. If no response occurred within 10 s, subjects received a third feedback display indicating that the trial had timed out. 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 that no specific set of instructions or examples explaining how to solve the problems were given to the subjects. Instructions on the computer screen prior to each block of test problems indicated the category of problems subjects were about to solve (e.g. 'addition in base 6').

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 and test session lasted about 1 h.

2.1.5. Data treatment

Latency data were analyzed for all trials on which a subject gave the correct response within the allotted 10 s time period. (Erroneous and timed-out trials were infrequent, and results are similar when all data are included.) For the training problems, mean response latencies were calculated separately for each subject, task, language, and training session.[2] For the test problems, mean response latencies were calculated separately for each subject, task, language, and problem type (trained or untrained). The principal analyses used parametric statistics based on the mean response latencies. Error rates were calculated for each subject, session, and condi-

[2] Because latency data are subject to positive skew, median reaction times also were calculated for each subject, session, and condition. Mean and median reaction times showed very similar patterns.

tion as the proportion of trials with either an erroneous or a timed-out response. Because these rates were low (see Figs. 1 and 2), they were not analyzed. Planned analyses tested (1) for the effect of monolingual versus bilingual number word training on performance on all problems, (2) for the effect of language change on performance on trained problems, and (3) for the effect of problem novelty on performance in the trained language. Further post-hoc analyses tested for other patterns in the data (see below).

2.2. Results

2.2.1. Training sessions

Fig. 1 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 types of problems in both languages. A 3 (task) × 2 (language of training: Russian versus English) × 2 (training day) ANOVA on the response latencies revealed main effects of training day ($F(1,7) = 183$, $P < 0.001$) and of task ($F(2,14) = 27$, $P < 0.001$), and 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 Tukey HSD tests, which indicated that subjects performed faster on the approximation tasks than on either the exact addition task with unfamiliar bases ($P < 0.05$) or the exact addition task with large numbers ($P < 0.01$), and faster on the unfamiliar bases task than on the large number addition task ($P < 0.01$). These effects may reflect both the greater length of the large number addition problems (on average those problems contained 86.2 characters, whereas the approximate problems and the novel base problems contained 68.3 and 61.7 characters, respectively) and the unfamiliarity of the novel bases task.

2.2.2. Test sessions

2.2.2.1. Effects of language on number word processing. We begin by comparing subjects' performance on all problems for which numbers were presented only in a single language to their performance on problems for which numbers were presented with equal frequency (in different problems) in both languages. Two 2 (number training: monolingual versus bilingual) × 2 (language of training: Russian versus English) × 3 (task) ANOVAs on the response latencies and response accuracy for all test problems presented in the untrained language revealed no main effects or interactions involving the number training variable. When bilingual subjects solved problems in a new language, they were no faster when the numbers in the problems had been trained in both languages (mean RT = 2840.2 ms) than when the numbers had been trained only in the language not being tested (mean RT = 2741.3 ms). Error rates also did not differ for bilingually versus monolingually-trained numbers (respective means 1.08 and 1.17%). Therefore, we disregard the number training factor in all further analyses.

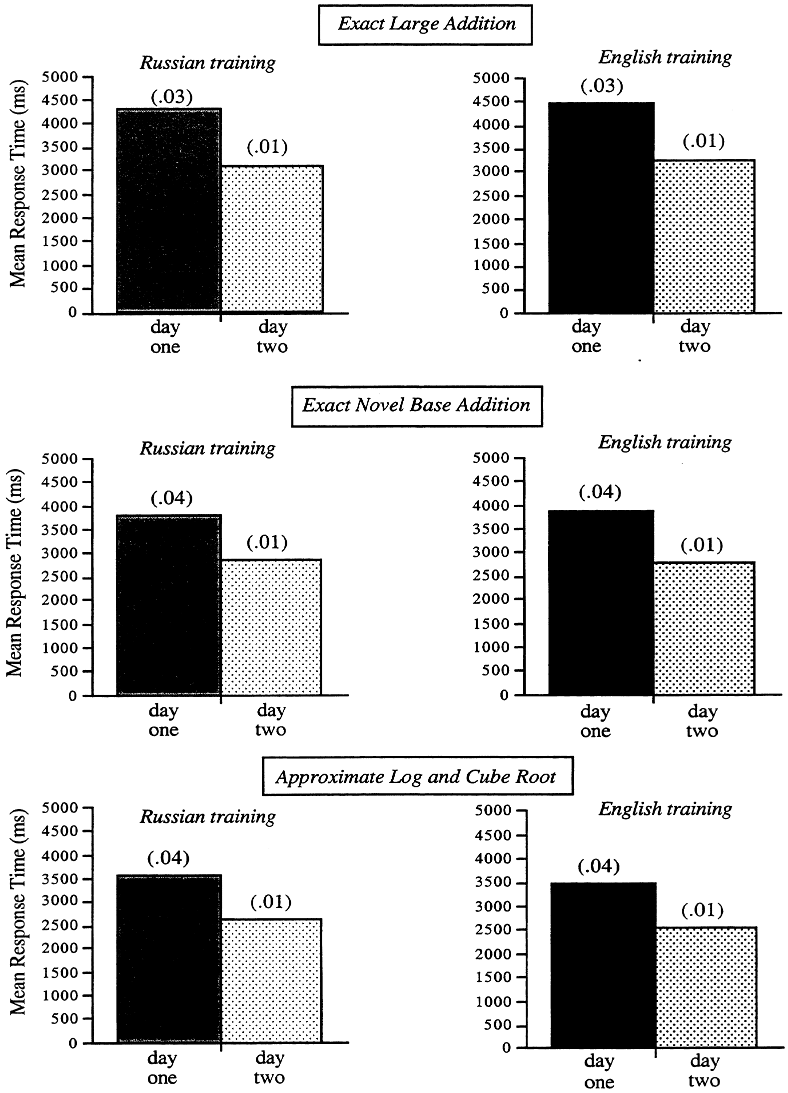

Fig. 1. Response latencies and error rates on each type of problem and in each language in Experiment 1 on the first and second days of training.

2.2.2.2. Generalization of performance on trained problems to a new language. Fig. 2 (left) 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 two tasks involving exact number representations, subjects were faster when performing in the language of training. This effect was obtained both when subjects were trained in their first language (Russian) and when they were trained in their

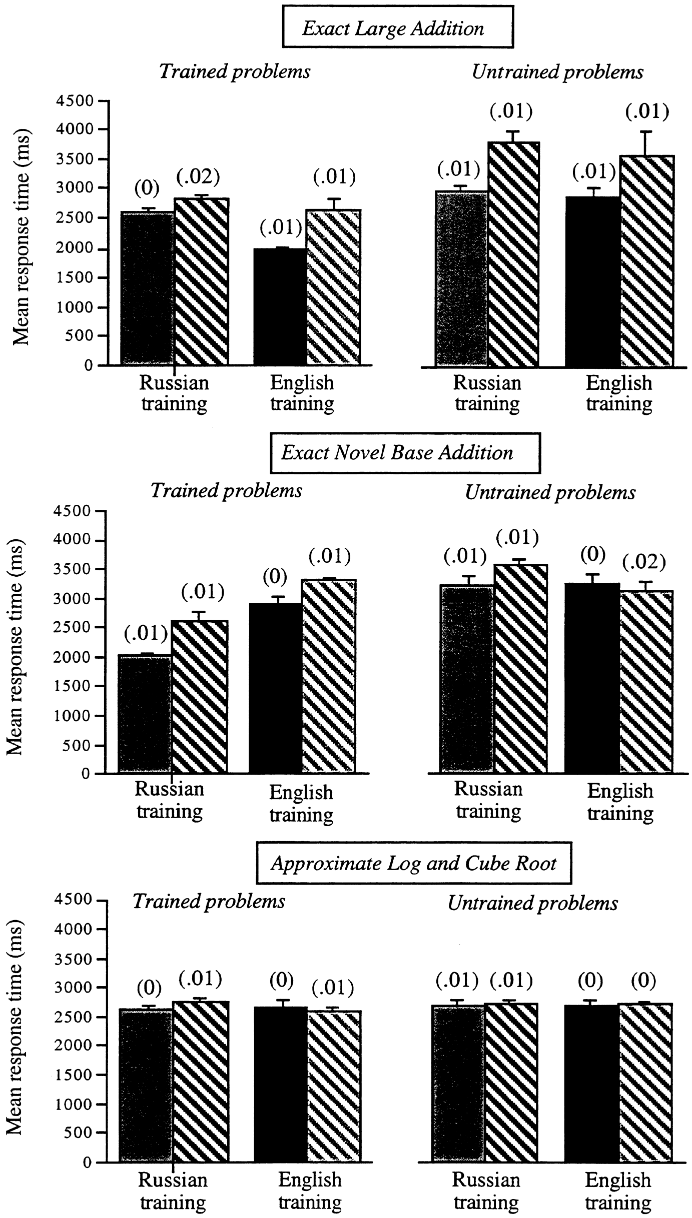

Fig. 2. Response latencies and error rates on each type of problem and in each language in Experiment 1 during testing in the trained and untrained languages (solid and stripped bars, respectively) on trained and untrained problems.

second language (English). For the tasks involving approximate number representations, in contrast, subjects performed with nearly equal speed in the trained and untrained languages. Accuracy was high on all tasks and in both languages.

For the latency data, a 3 (task) × 2 (language of training: Russian versus English) × 2 (language change: same language as training versus different language) ANOVA revealed main effects of task ($F(2, 14) = 4.67$, $P < 0.05$) and language change ($F(1, 7) = 48.8$, $P < 0.001$), and an interaction between these factors ($F(2, 14) = 17.8$, $P < 0.001$). There was also an interaction between task and language of training ($F(2, 14) = 30.8$, $P < 0.001$). Tukey tests revealed that subjects were faster overall at solving the exact large addition problems than at solving the exact addition problems with novel bases ($P < 0.05$). Moreover, subjects were faster in solving both types of exact addition problems in the language of training (each $P < 0.01$) but solved the approximation problems with equal speed in the two languages. Finally, subjects performed the exact large additions faster after training in English than after training in Russian ($P < 0.01$), and they performed the exact additions with novel bases faster after training in Russian than after training in English ($P < 0.01$).

To understand these interactions better, separate analyses were performed for each of the three types of task. For the exact large addition tasks, the 2 (language of training) × 2 (language change) ANOVA performed on the latency data revealed main effects of language of training ($F(1, 7) = 18.9$, $P < 0.005$) and of language change ($F(1, 7) = 24.2$, $P < 0.005$), 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. In neither case was greater speed accompanied by lower accuracy (see Fig. 2).

For the exact addition tasks with novel bases, the same 2 × 2 ANOVA revealed the same main effects of language of training ($F(1, 7) = 29.5$, $P < 0.005$) and language change ($F(1, 7) = 304.1$, $P < 0.001$), and no interaction. Subjects performed faster after training in Russian, and they performed faster when answering questions in the trained language. Accuracy levels suggested that a speed–accuracy trade-off had little influence on either of these effects (Fig. 2).

For the approximate log and cube root tasks, in contrast, the same 2 × 2 ANOVA revealed no main effects or interactions (all $F < 1.2$). Subjects responded with equal speed after training in English and Russian, and they responded with equal speed when tested in the trained and in the untrained language.

2.2.2.3. Generalization of performance in the trained language to new problems. Turning now to the question of whether training generalized to new problems, Fig. 2 (right) presents the mean accuracy and response latencies of performance on the novel problems for each language and task. A comparison of the solid bars on the left and right sides of Fig. 2 reveals a difference between the exact and the approximate number tasks. For the approximate tasks, untrained and trained problems were solved with equal speed, providing evidence that training generalized beyond the particular problems that subjects learned. For the exact tasks, in contrast, untrained problems were solved more slowly than trained problems for both languages of training. Accuracy was high for both old and new problems. The differences in response latencies were confirmed by a 3 (task) × 2 (language of training: Russian versus English) × 2 (problem novelty: trained versus untrained problems) ANOVA on response latencies to test problems presented in the language of training. The analysis revealed a main effect of problem novelty ($F(1,7) = 89.4$, $P < 0.001$), qualified by interactions between problem novelty and task ($F(2,14) = 15.6$, $P < 0.001$), and problem novelty, task and language of training ($F(2,14) = 30.2$, $P < 0.001$). In addition, there was again an interaction of task and language of training ($F(2,14) = 12.8$, $P < 0.005$). Tukey HSD tests applied to the interactions revealed that the subjects were faster at solving old than new exact addition problems both in base 10 and in unfamiliar bases (both $P < 0.01$), but they were equally fast at solving old and new approximate log and cube root problems. Overall, subjects were faster at solving both old and new exact large addition problems when trained and tested in English ($P < 0.05$), and they were faster at solving both old and new exact addition problems in novel bases when trained and tested in Russian ($P < 0.01$).

To investigate the interactions further, separate analyses of the problem novelty effect were performed for each task. For the exact large addition task, the 2 (language of training) × 2 (problem novelty) ANOVA revealed main effects of language of training ($F(1,7) = 6.83$, $P < 0.05$) and problem novelty ($F(1,7) = 28.61$, $P < 0.001$), and an interaction between these factors ($F(1,7) = 14.52$, $P < 0.01$). Subjects performed faster in English than in Russian, they performed faster on trained than on untrained problems, and the advantage for trained problems was larger in English. For the exact novel bases task, the ANOVA revealed significant effects of language of training ($F(1,7) = 73.74$, $P < 0.001$) and problem novelty ($F(1,7) = 71.10$, $P < 0.001$), and an interaction between the factors ($F(1,7) = 25.17$, $P < 0.005$). Subjects performed faster in Russian than in English, they performed faster on trained than on untrained problems, and the advantage for trained problems was larger in Russian. For the approximate logs/cubes task, the ANOVA revealed no significant effects (all $F < 1$).

2.2.2.4. Further effects. Inspection of the data in Fig. 4 suggests two further effects of training on subjects' performance. First, subjects appeared to perform better on new exact problems when they were tested in the language in which they learned similar problems (i.e. problems sharing an addend with the test problems for the exact large addition task and problems in the same base for the exact novel bases task) than when they were tested in the language in which they learned less similar problems (i.e. problems with two different addends or in a different base). This apparent effect was analyzed by a 3 (task) × 2 (language of training) × 2 (language change) ANOVA, which revealed significant main effects of task ($F(2,14) = 15.28$, $P < 0.001$) and language change ($F(1,7) = 22.34$, $P < 0.005$), and an interaction between these factors ($F(2,14) = 7.55$, $P < 0.01$). Separate analyses of each of the three tasks revealed that the effect of language change (i.e. training with similar problems in the same versus different language) was significant for the exact large addition task ($F(1,7) = 15.4$, $P < 0.01$), but not for the exact novel bases task or for the approximate logs/cubes task (both $F < 1$).

Second, when subjects solved exact number problems (both large number addition and addition in novel bases) in the untrained language, they appeared to perform faster on problems on which they had been trained in their other language than on new problems. This effect was confirmed by a 3 (task) × 2 (training language) × 2 (problem novelty) ANOVA on response latencies to test problems presented in the untrained language. The analysis revealed main effects of task ($F(2, 14) = 22.4$, $P < 0.001$) and problem novelty ($F(1, 7) = 32.8$, $P < 0.001$), and interactions between task and problem novelty ($F(2, 14) = 10.3$, $P < 0.005$) and task, language of training and problem novelty ($F(2, 14) = 9.0$, $P < 0.005$). A separate ANOVA on the exact large addition task revealed a main effect of problem novelty ($F(1, 7) = 17.11$, $P < 0.005$) and no other effects. A separate ANOVA on the exact novel bases task revealed the same main effect of problem novelty ($F(1, 7) = 38.41$, $P < 0.001$), and an interaction of problem novelty with language of training ($F(1, 7) = 13.26$, $P < 0.01$), indicating that the problem novelty effect was greater when the untrained language was English. Finally, a separate ANOVA on the approximate tasks revealed no effect of problem novelty ($F < 1$) and no other significant effects.

2.3. Discussion

Bilingual subjects who learned new facts about approximate numbers in one language 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 on representations that are independent of language.

In contrast, when the same subjects learned new facts about exact numbers in one language, 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 exact number in one of the two languages, because it was observed in each subject for both the subject's languages. 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 the training sessions on the efficiency of translation processes between the number words and a language-independent representation, because it was as large for problems involving bilingually-trained number words as for those involving monolingually-trained number words. These findings provide evidence that exact number facts, involving both the familiar base 10 addition and the more novel addition operation in bases 6 and 8, are represented at least partly in a language-specific form, in accord with the theses of Dehaene (1997) and Gallistel and Gelman (1992) and contrary to that of McCloskey (1992).

The students' performance in solving new problems within the training sets (e.g. calculating the cube root of a new four-digit number, or adding 54 to a new two-digit number) sheds further light on the nature of their number representations. For the log and cube root estimation problems, learning generalized fully from old to new problems in both languages. This finding suggests that subjects represented each large number as 'a blur on the number line' (Gallistel & Gelman, 1992; see also Dehaene, 1997). Because each approximate numerosity was represented as a range of numbers rather than as a single discrete point, subjects' learning generalized to other facts within the range. The approximate number representations formed by these bilingual subjects therefore appear to be similar to those observed in a variety of animals, as others have proposed (Dehaene, 1997; Gallistel, 1990).

For the exact number tasks, three findings emerged from the analyses of performance on new problems. First, new problems were solved in the trained language more slowly (and with equal or higher error rates) than old problems. This finding provides evidence that subjects represented the numbers in the exact addition problems in a manner that failed to generalize to other, neighboring numbers. Indeed, subjects may have represented each exact addition fact as a string of words.

Second, subjects solved new exact large addition problems more effectively when they were tested in a language in which they had been trained on similar problems (i.e. problems sharing one of the addends) than when they were tested in a language in which they had been trained on less similar problems (problems with two different addends). No such effect of language change was observed for the exact novel bases problems (in which the addends were equally likely to differ across languages) or for the approximation problems. This effect suggests that language-specific learning influences exact arithmetic performance not only when one confronts exactly the same problem on which one was trained but also when one confronts a new problem that shares features of the trained problem.

Third, subjects showed an advantage for solving trained over untrained problems in the untrained as well as the trained language. In the untrained language, subjects' more efficient performance with trained problems might be taken to suggest that the representations underlying exact number arithmetic learning are not entirely language-specific. Exact large number representations may involve both language-dependent and language-independent processes (Dehaene, 1997; Gallistel & Gelman, 1992). Alternatively, learning exact number addition facts may be entirely language-dependent, but subjects may use facts learned in one language to solve problems in their other language by translating problems into the language of training. On the latter view, subjects may perform more quickly with trained problems in the untrained language because translating the 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 have suggested that performance on the log and cube root problems differed from performance on the addition problems because the former involved approximate number representations, whereas the latter involved exact number representations. Nevertheless, the findings of Experiment 1 are consistent with other possible distinctions between language-dependent and language-independent learning. For example, it is possible that facts involving binary operations are learned in a language-specific manner, whereas facts involving unary operations are not. As a second example, it is possible that facts with two- to three-digit answers or with unfamiliar novel base answers depend on language, whereas facts with familiar small number answers do not. Recall that subjects answered the log and cube root problems more quickly during training than the exact addition problems. Although this effect could stem from factors such as the differing lengths of the different verbal problems and the differing familiarity of the notation used in different tasks (see above), it could also be explained by differences in the processing of unary versus binary operations or of highly familiar versus less familiar numbers. To distinguish these possibilities, the next experiment contrasted subjects' performance of a single arithmetic operation – addition – under conditions requiring either exact or approximate representations of number.

3. Experiment 2

The primary purpose of Experiment 2 was to investigate whether exact, but not approximate, number representations are language-dependent, by comparing Russian–English bilingual subjects' performance on base 10 addition when either an exact or an approximate answer was required. All subjects were taught a set of new large number addition facts in one language and then were tested on knowledge of those facts in both their languages. Subjects were trained under either of two conditions. One group of subjects learned the exact answers to these facts, and the other group learned the approximate answers to these facts. The first condition was essentially a replication of the exact large arithmetic task in Experiment 1; as in that experiment, subjects 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 language-specific processing is engaged by binary operations such as addition or by processing of large or unfamiliar numbers, then subjects taught new approximate number addition facts should show the same advantage for the language of training as those taught exact number addition facts. In contrast, if language-specific processing is engaged by the representation of exact large numbers, then the subjects taught new approximate number facts should perform equally well in their two languages, as did subjects taught new log and cube root estimation problems in Experiment 1.

Experiment 2 had two subsidiary purposes. First, it investigated further the finding from Experiment 1 that learning to estimate the answers to arithmetic problems generalizes to new problems of the same type, whereas learning the exact answers to problems results in better performance of trained than of untrained problems in both the trained and the untrained language. 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.

A further 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 approximate, language-independent number representations found in animals serve as inputs to the operation of multiplication. The evidence for multiplication processes in animals, however, is indirect and open to other explanations (e.g. Church & Meck, 1984). Based on analyses of human patient data and contrasting analyses of animal experiments, Dehaene (1997) proposed that approximate number representations enter into the operations of addition and subtraction but not multiplication. To shed further light on this controversy, each subject in Experiment 2 was taught a set of new multiplication facts in the language not used for the addition training. Subjects trained on exact addition facts were taught new approximate multiplication facts, whereas those trained on approximate addition facts were taught new exact multiplication facts.[3] If humans have a language-independent system for multiplying approximate large numerosities, then the subjects taught approximate number multiplication facts should generalize their learning both to the untrained language and to untrained problems involving numbers within the same range. If no language-independent multiplication system exists, in contrast, then all the subjects should have shown language-specific and item-specific learning of multiplication facts, regardless of whether exact or approximate answers were required.

3.1. Method

The method was the same as for Experiment 1, except as follows.

3.1.1. Subjects

The participants were three female and five male bilingual speakers of Russian and English, ranging in age from 18 to 32 years (mean 22.5 years). Subjects were undergraduate or graduate students at Cornell University who began learning English at a mean age of 15.3 years (range 12–18 years) and who had been in the US for an average of 4.9 years (range 3.5–6.5 years). All subjects passed the preliminary comprehension test from Experiment 1, and all spoke and comprehended both Russian and English with ease.

3.1.2. Materials

Subjects were tested with the same computer as in Experiment 1, with Adobe Times Ten Cyrillic font used for the Russian stimuli. The problem sets were as follows (see Appendix A for examples):

(1) *Exact addition*: 12 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 distractor in which the tens place was increased or decreased by 1.

(2) *Approximate addition*: these problems were identical to the exact addition problems except for the candidate answers, which ranged from 50 to 160. All

[3] Training on multiplication facts served two further purposes: it assured that subjects were equally familiar with the general experimental procedures in both their languages prior to the test, and it allowed us to equate subjects' exposure to specific number words in their two languages.

response alternatives were multiples of 10; one was the nearest such multiple and the other was 30 units larger or smaller than that multiple.

(3) *Exact multiplication*: 12 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 10 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 10; one was the nearest such multiple and the other was 30 units larger or smaller than that multiple.

3.1.3. Design

Each subject participated in four training sessions and two test sessions. A training session consisted of one set of problems in one language; every subject was trained on two sets of problems, one in each language, for two sessions each. Half the subjects were trained on exact addition and approximate multiplication problems, and half were trained on approximate addition and exact multiplication problems. A test session consisted of blocked presentations of all four sets of problems in one language, 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, each problem set consisted of 12 different problems presented six times (72 problems per set and per session). During testing, a problem set consisted of six problems that the subjects were trained on and six new problems of the same type. Each problem was presented twice during the course of testing (24 problems per set; 96 problems per session). Every blocked set of problems was preceded by three filler problems, after which problems appeared in a random order with the restriction that no problem could appear twice in succession. Within a set, the correct answer appeared on the left and right with equal frequency.

3.1.4. Procedure

Before the study, the exact and the approximate tasks were each explained 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 response closest to it, but rather to estimate the answer directly. Because training sessions were short (about 20 min), subjects were given two sessions per day separated by a 10 min break. Each test session lasted about 45 min.

3.2. Results

3.2.1. Addition training

Fig. 3 presents the mean response latencies and the error rates 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 in both languages. A 2 (task: exact versus approximate addition) × 2 (language:

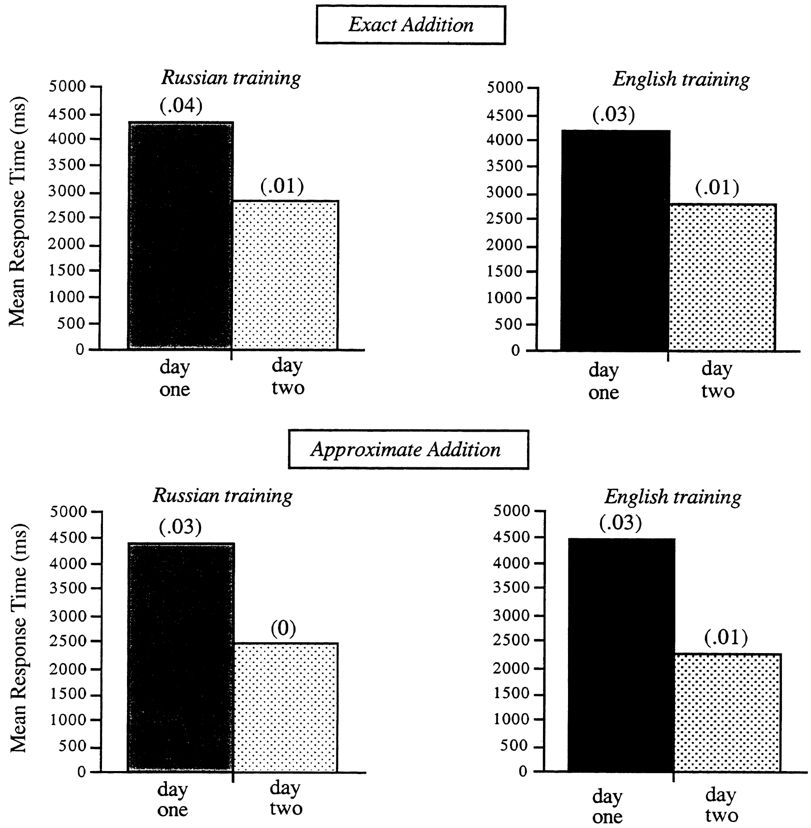

Fig. 3. Response latencies and error rates on exact and approximate addition problems in each language in Experiment 2 on the first and second days of training.

Russian versus English) × 2 (session) ANOVA on the response latencies revealed only a main effect of session ($F(1,4) = 297.2$, $P < 0.001$), and an interaction of session with task ($F(1,4) = 8.1$, $P < 0.05$). Improvement in performance was greater for the approximate addition task than for the exact addition task.

3.2.2. Addition testing

Fig. 4 (left) presents the mean response latencies and the error rates for the trained addition problems tested in both languages. Subjects who were trained on exact addition problems performed faster 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 answered those problems with nearly equal speed in the trained and untrained languages. This difference was not attributable to a speed–accuracy trade-off, because accuracy rates were high and slightly favored the language of training for both exact and approximate problems. The latency findings were confirmed by a 2 (task: exact

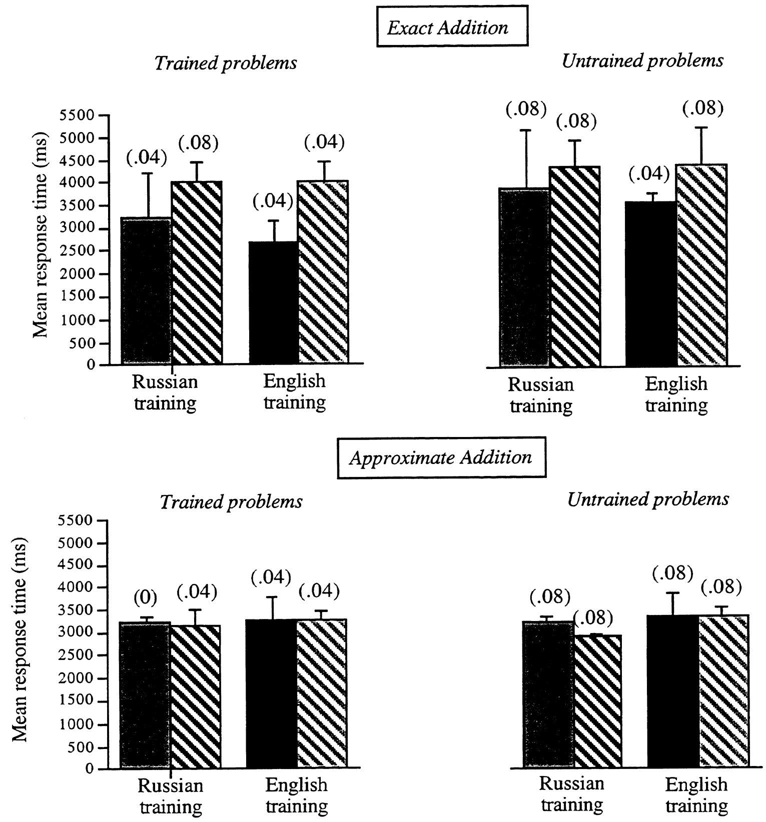

Fig. 4. Response latencies and error rates on exact and approximate addition problems in each language in Experiment 2 during testing in the trained and untrained languages (solid and striped bars, respectively) on trained and untrained problems.

versus approximate addition) $\times$ 2 (training language: Russian versus English) $\times$ 2 (language change: test in trained versus untrained language) ANOVA which revealed a main effect of language change ($F(1,4) = 19.8$, $P < 0.025$), qualified by an interaction between task and language change ($F(1,4) = 22.0$, $P < 0.01$). Subjects performed better when tested in the language of training for the exact addition problems but not for the approximate addition problems.

Fig. 4 (right) presents subjects' responses on the new exact and approximate addition problems, both in the language of training and in the untrained language. A comparison of the solid bars on the left and right sides of Fig. 4 indicates that subjects trained at approximate addition answered old and new problems with equal speed in the trained language, whereas those trained at exact addition answered old problems more rapidly than new problems in that language. Responses on old problems were more accurate in both conditions. These effects were analyzed first by a 2 (task) $\times$ 2 (training language) $\times$ 2 (problem novelty: old versus new problem)

ANOVA on response latencies for test problems performed in the language of training. The analysis revealed a main effect of problem novelty ($F(1,4) = 9.9$, $P < 0.05$) and a borderline-significant interaction of problem novelty with task ($F(1,4) = 7.6$, $P = 0.051$). When tested in the language of training, subjects responded faster when solving old than new exact addition problems, but they responded with equal speed when solving old and new approximate addition problems.

Inspection of Fig. 4 suggests two further effects in Experiment 2, parallel to those from Experiment 1. First, subjects appeared to perform better on exact untrained problems when tested in the language in which they had learned similar problems (i.e. other exact addition problems) than when tested in the language in which they had learned different problems (i.e. approximate multiplication problems); no such effect appeared for performance on approximate untrained problems. This pattern, however, was found not to be significant. A 2 (task) × 2 (training language) × 2 (language change) ANOVA on response latencies for novel test problems revealed no significant main effects or interactions (all $P > 0.10$). Second, subjects appeared to perform better on old than on new problems when tested in the untrained language. A 2 (task) × 2 (training language) × 2 (problem novelty) ANOVA on response latencies for test problems performed in the untrained language revealed a main effect of task ($F(1,4) = 11.8$, $P < 0.05$), qualified by an interaction between task and problem novelty ($F(1,4) = 7.8$, $P < 0.05$). In the untrained language, old exact addition problems were answered faster than new exact addition problems, whereas old and new approximate addition problems were answered with equal speed.

3.3. Results: multiplication tasks

3.3.1. Training sessions

Fig. 5 presents the mean response latencies and error rates for the exact versus approximate multiplication tasks in each language and session. Speed and accuracy improved from the first to the second session for both tasks and both languages, although performance was faster initially for the approximate problems and the improvement appeared to be steeper for the exact problems. These effects were confirmed by a 2 (task: exact versus approximate multiplication) × 2 (language of training) × 2 (session) ANOVA which revealed significant main effects of session ($F(1,4) = 164$, $P < 0.001$) and task ($F(1,4) = 14.1$, $P < 0.025$), and an interaction of these factors ($F(1,4) = 13.3$, $P < 0.025$).

3.3.2. Testing sessions

Fig. 6 present the response latencies and error rates for the trained and untrained multiplication problems in both the trained and the untrained languages. For both trained and untrained problems, subjects performed better on approximate than on exact problems and they performed better in the language of training than in the untrained language. A 2 (task: exact versus approximate multiplication) × 2 (language of training: Russian versus English) × 2 (language change: same versus

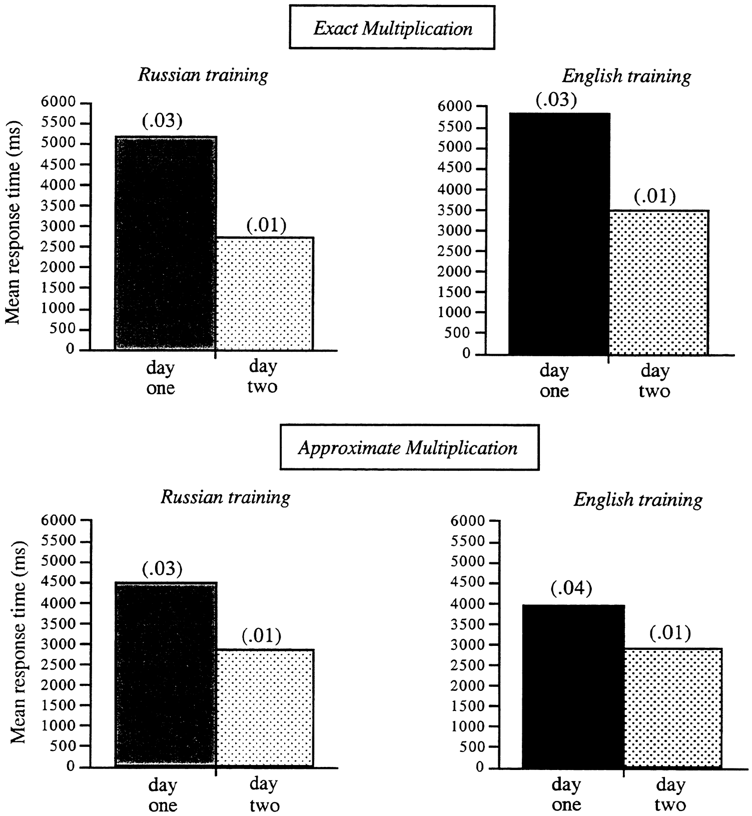

Fig. 5. Response latencies and error rates on exact and approximate multiplication problems in each language in Experiment 2 on the first and second days of training.

different) × 2 (problem novelty: trained versus untrained) ANOVA revealed main effects of task ($F(1,4) = 65.16$, $P < 0.001$) and language change ($F(1,4) = 12.14$, $P < 0.05$). There was no interaction of task by language change ($F < 1$). Exact multiplication showed no greater language dependence than approximate multiplication. In general, performance on the multiplication task showed greater variability than the addition task. The trend toward better performance on trained than on untrained problems was not significant in the above analysis ($F(1,4) = 2.21$, $P > 0.2$), and the effect of language change was not significant when the data from the trained problems and untrained problems were analyzed separately ($F = 2.05$ and 2.15, respectively, $P < 0.2$). In addition, performance on the approximate multiplication tasks showed high error rates, complicating interpretation of the latency data.

3.4. Discussion

The principal findings of Experiment 2 confirm and extend those of Experiment 1.

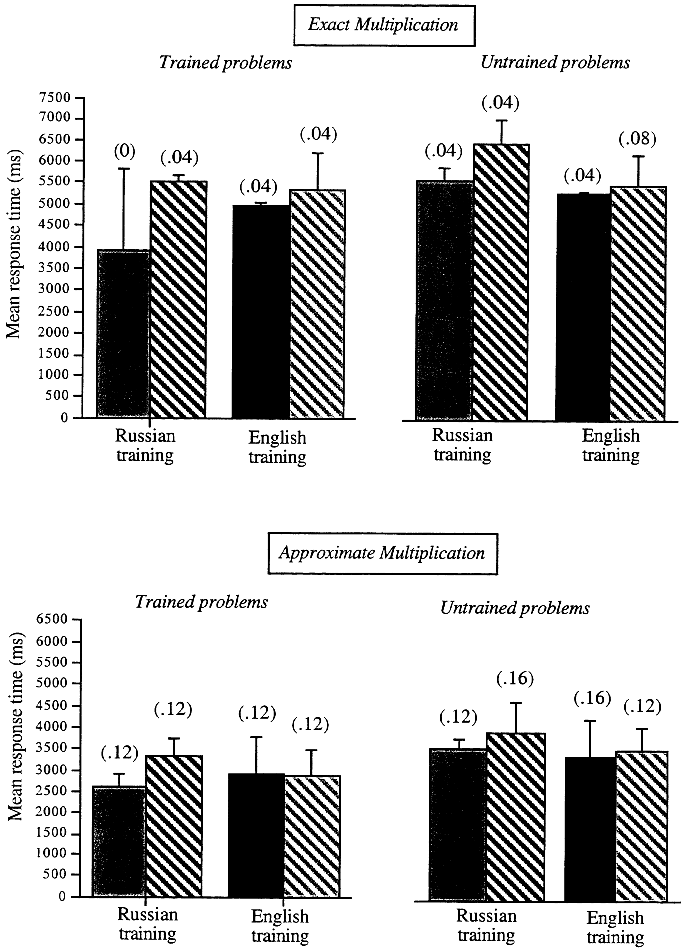

Fig. 6. Response latencies and error rates on exact and approximate multiplication problems in each language in Experiment 2 during testing in the trained and untrained languages (solid and striped bars, respectively) on trained and untrained problems.

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 adds 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. The critical difference between language-dependent and language-independent arithmetic tasks stems not from the nature of the operations performed (unary versus binary) or from the range of the numbers in the answers (small versus large) but from the nature of the number representations required. Representations of approximate numerosity are independent of language, whereas representations of exact numerosity are language-specific.

Experiment 2 also replicated the other principal finding of 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 involving numbers within the same range. Experiment 2 therefore provides further evidence that learning new facts about approximate numerosities invokes representations of those numerosities as 'blurs' on the number line, allowing generalization to neighboring numerosities. In contrast, learning new facts about exact numerosities invokes different representations that do not privilege connections among neighboring numbers.

One may ask whether the different patterns of generalization observed for exact versus approximate arithmetic tasks stem from differences in subjects' learning and memory strategies rather than from differences in their number representations. Perhaps subjects who perform exact arithmetic come to store and retrieve the answers to the problems they solve, whereas subjects who perform approximate arithmetic compute the answer anew every time a problem is presented. If subjects failed to store or retrieve the approximate answers to problems, then obviously they would perform no better on trained problems than on new problems, and they would show no language-of-training advantage. Subjects' equal performance on all approximate problems in both languages therefore would not reflect true generalization but rather the absence of true learning.[4]

Subjects' performance in the training sessions provides evidence against this alternative interpretation of their generalization performance. If subjects stored exact but not approximate answers to the trained problems, then they should have shown a greater benefit from training on the exact problems than on the approximate problems, i.e. a greater decrease in response times or error rates. In fact, the training benefit for the approximate problems was as large as that for the exact problems in Experiment 1, and it was larger than for the exact problems in Experiment 2. These findings provide evidence that subjects learned and retrieved as much or more information about the approximate problems as about the exact problems. The different patterns of language specificity and problem specificity observed on the test reflect true generalization of learning, rather than the absence of learning. This generalization, in turn, provides evidence for differences in the representations of number used in exact versus approximate arithmetic tasks.

Like Experiment 1, Experiment 2 provided evidence that subjects solved old exact large addition problems faster than new problems when tested in the untrained as well the trained language. Again, it is not clear whether subjects' performance on old problems was fostered by rapid translation processes from the trained

[4] We thank an anonymous reviewer for raising this possibility.

language or from an effect of training on language-independent representations. In contrast to Experiment 1, subjects performed no better on novel exact large addition problems when they were tested in a language in which they had learned similar problems than when they were tested in a language in which they had learned different problems. The absence of an effect of language change on novel exact addition problems in Experiment 2 likely stems from the fact that in Experiment 2, in contrast to Experiment 1, new addition problems did not share an addend with old addition problems.

Finally, Experiment 2 provides little evidence for any language-independent representation of multiplication facts. Performance in the language of training exceeded performance in the untrained language, and this effect of language change was no greater on the exact than on the approximate problems. These findings support Dehaene's (1997) thesis that non-verbal number representations are accessible to the addition operation but not to multiplication. If subjects cannot solve multiplication problems through a language-independent representation of approximate numerosities, then 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 than on the exact multiplication problems, because subjects do not need to complete the full multiplication to choose the closer answer. The same language-dependent representations of exact number nevertheless may underlie performance on both the exact and the approximate multiplication problems. No strong conclusions about the mental representation of multiplication facts can be drawn from the present experiment, however, because of the high variability and high error rates shown in the multiplication tasks.

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 material outside any arithmetic context: simulated lessons in history and geography.

4. Experiment 3

The idea for these experiments began when the first author discovered that she could readily provide American friends with her summer address in France but not with her telephone number. Retrieving the number required that she say it in (non-native) French, visualize the numerals, and then mentally read them off in English. Pursuing the anecdote, she discovered that her (American) children could tell her rapidly, in English, the occasion of France's independence day ('the taking of the Bastille') but not its date ('July…err…le quatorze juillet!'). Are exact numbers always stored in a language-dependent form, even when they do not appear in arithmetic expressions? Our final bilingual training study begins to address this question.

In this study, Russian–English bilingual students learned a fictitious history lesson in one of their languages and a fictitious geography lesson in the other language. Over the course of three training sessions for each lesson, subjects learned and were tested on material involving large numbers and material involving non-numerical information. For example, one lesson specified that 'Almost four hundred thirty years ago in a country known as Kapnopa, a band of farmers met in a secret, underground cave below the marketplace', and subjects were asked both when the farmers met (alternative answers: 430 versus 480 years ago) and where they met (alternative answers: in a cave versus in a house). The primary purpose of Experiment 3 was to test whether subjects learned the numerical and non-numerical material in a language-specific or language-independent manner. To this end, subjects were subsequently tested, during one session in Russian and one session in English, on both numerical and non-numerical material in both stories, so as to compare their speed and accuracy of responding in the language of training and in the untrained language.

Experiment 3 also tested further whether approximate numerical information is stored independently of language when subjects deliberately attempt to learn both numerical and non-numerical information exactly. If language-independent, approximate number representations form automatically during explicit learning of exact numerical facts, then new facts about exact numbers that are very small (below 5) should be equally accessible to them in the language of training and in the untrained language.

To test this prediction, a small set of facts containing the ordinal numbers between 1 and 4 (e.g. 'the second election') and containing simple fractions (e.g. 'three-fourths of all recorded ancient legends') was embedded in the lessons and tested in both languages. Because the discriminability of approximate numerosity is proportional to set size, in accord with Weber's Law (see Gallistel, 1990), we reasoned that language-independent, approximate number representations should be sufficient to capture these small numerosities exactly. If subjects automatically formed language-independent, approximate number representations of these facts, therefore, responses to these questions should have been equally fast and accurate in the trained and untrained languages.

The final purpose of Experiment 3 was to begin to probe whether other categories of information, besides exact number, are storied in a language-specific or language-independent manner. It has been proposed that humans form language-dependent representations of egocentric spatial directions (Hermer & Spelke, 1996; Hermer-Vazquez, Spelke, & Katsnelson, 1999), geocentric spatial directions (Levinson, 1996), and time (Peacocke, 1992). As a preliminary test of these proposals, small amounts of information in each of these categories were 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.

4.1. Method

The method was the same as in Experiments 1 and 2, except as follows.

4.1.1. Subjects

The participants were six female and two male bilingual speakers of Russian and English, ranging in age from 19 to 33 years (mean 24 years). One additional subject was dropped from the study when the screening test revealed poor English comprehension (see below). All the retained subjects were undergraduate or graduate students living in the greater Boston, MA area and were solicited through an ethnic club at MIT and through posted ads on the campuses of MIT and neighboring universities. The mean age at which subjects started learning English was 16 years (range 13–18 years), and the mean time since coming to the US was 5 years (range 3–7 years). All subjects were perceived by the experimenter to speak and comprehend Russian and English with ease.

4.1.2. Materials

Training was conducted with written lessons printed single-spaced on two pages of 8.5 × 11 inch paper, with a count of 3712 characters, 623 words and 52 lines for the history lesson in English, 3262 characters, 528 words and 75 lines for the history lesson in Russian, 4480 characters, 769 words and 60 lines for the geography lesson in English, and 3763 characters, 677 words and 85 lines for the geography lesson in Russian. The lessons were printed in English in Times font size 14, and they were printed in Russian in Times Ten Cyrillic font of the same point size, 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; the geography lesson detailed the travels and adventures of a fictitious character. In both lessons, numbers were written out as words in the appropriate language, rather than as Arabic digits.

All testing of comprehension and retention took place on a Power PC Macintosh computer with a 17 inch screen. Each question was presented on the monitor with a picture size of 512 × 600 pixels. 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 Adobe Times Ten Cyrillic in the same point sizes as for the English materials.

For each of the lessons, questions probed subjects' comprehension and memory for information in six categories as follows (see Appendix A for examples):

(1) *Exact large numbers*: subjects were asked about 16 large number facts in each lesson, such as the age of a character, the duration of an event, or the length of a journey. Half of the facts used numbers that also appeared in the other lesson in a different context ('bilingually-trained numbers'), and half used numbers appearing only in a single lesson ('monolingually-trained numbers'). The tested numbers 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 lesson; for the remaining questions, the distractor answer did not

appear in the lesson. The distractor always differed from the correct answer in either the units or the tens place by any amount between 1 and 9.

(2) *Object categories*: subjects were asked 16 questions involving no numbers, whose answers were the names of common objects. For half of the questions in each lesson, the correct answer and distractor also appeared as the two alternative answers for a question in the other lesson ('bilingually-trained words'); for the remaining questions, answers appeared only in one lesson. For half the questions, the distractors were words drawn from the same lesson; for the remaining questions, the distractors were words that did not appear in that lesson.

(3) *Exact small numbers*: subjects were asked about four facts involving fractions with a numerator and a denominator under 5 and four facts involving the first four ordinal numbers, for a total of eight small number questions. All distractors were chosen from the same set of fractions and ordinal numbers as the correct answers, and all answers appeared in both lessons.

(4) *Spatial relations*: subjects were asked 12 questions probing the retention of information concerning spatial directions. Six questions probed the learning of egocentric directions, with the candidate answers of 'left', 'right', 'back', 'front', 'beside' and 'behind'. Four questions probed the learning of geocentric directions, with the candidate answers of 'west', 'east', 'south' and 'north'. The remaining two questions (with the candidate answers of 'above' and 'below') could represent either egocentric or geocentric spatial directions. Distractors were drawn from the same set of spatial terms as the correct answers. All candidate answers appeared in both lessons.

(5) *Times of day and year*: subjects were asked questions concerning the time of day (e.g. 'mid-morning') or the season of the year (e.g. 'winter') when a target event took place. For the four time of day questions, distractor terms did not appear in either lesson; for the four season questions, distractors appeared in the same lesson. The same terms were queried in both lessons.

(6) *Proper names*: six questions concerning the names of characters and landmarks were used as fillers for each lesson and were not analyzed. None of the queried names appeared in more than one lesson; all distractors were names that appeared in the same lesson.

Before the study, subjects were given comprehension tests in Russian and English. Subjects first were 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 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 one (who was dropped from the experiment for failure to name the precious stones in English) correctly named all the key target items in both languages.

4.1.3. 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 lessons were orthogonally counterbalanced across sessions. During each session, subjects studied one lesson and answered 12 questions. Over the course of the training sessions, subjects encountered a total of 36 different questions per lesson, of which eight tested large number knowledge, eight tested non-numerical knowledge, four tested small number knowledge, six tested spatial knowledge, four tested temporal knowledge, and six 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 lesson.

After training, subjects received one test session in each language. Each test session consisted of 60 questions from the history lesson and 60 questions from the geography lesson. Each set of 60 questions consisted of 16 questions testing large number knowledge, 16 questions testing non-numerical knowledge, eight questions testing small number knowledge, 12 questions testing spatial knowledge, and eight questions testing temporal knowledge. Half the questions from each category for each lesson had been presented during training ('old questions') and half had never previously been presented. Each question was asked twice, for a total of 240 questions/session.

4.1.4. Procedure

At the start of each training session, subjects were given general instructions in the language to be used for that session, and then they were presented with one lesson. The lesson 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 lesson was followed by two questions (presented in random order), 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 lesson. At the start of each test session, subjects were told, in the language appropriate for that session, that they would be reading and answering questions about both of the lessons that they had studied. Subjects were not given the lessons to read again. Instead, they were given the 60 test questions in random order, followed by a repeat of the same 60 questions in a different random order. Breaks were taken after every set of 30 questions.

4.2. Results

4.2.1. Training sessions

Fig. 7 presents the mean reaction times and error rates for questions on the lessons presented in each language and on each of the days of training. Accuracy increased

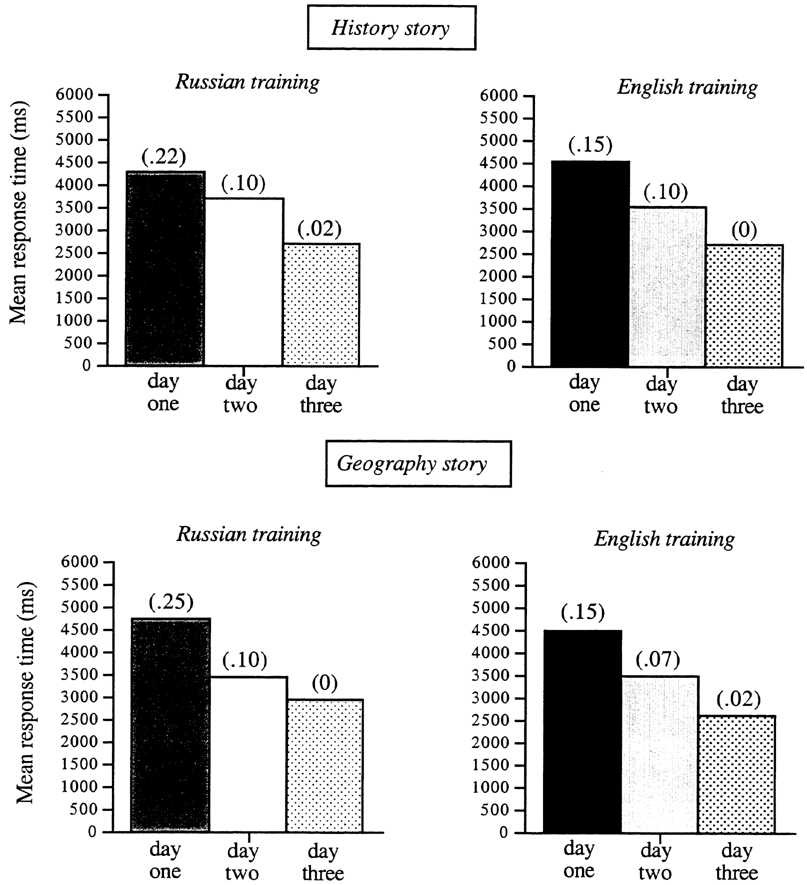

Fig. 7. Response latencies and error rates on each type of problem and in each language in Experiment 3 on each of 3 days of training.

and latency declined from day 1 to day 3 for the questions in both languages; performance was similar for the two languages and the two lessons. A 2 (lesson: history versus geography) × 2 (training language: Russian versus English) × 3 (training day) ANOVA on the response latencies revealed only a main effect of training day ($F(2,6) = 107.9$, $P < 0.001$). A Tukey test indicated that the subjects performed faster on day 3 than on days 1 or 2, and faster on day 2 than on day 1 (all $P < 0.01$).

4.2.2. Testing sessions

Fig. 8 presents the mean reaction times and error rates for the large number fact questions and for the non-numerical questions, both in the language of training and in the untrained language. For the large number facts, subjects answered more rapidly in the language of training than in the untrained language, regardless of

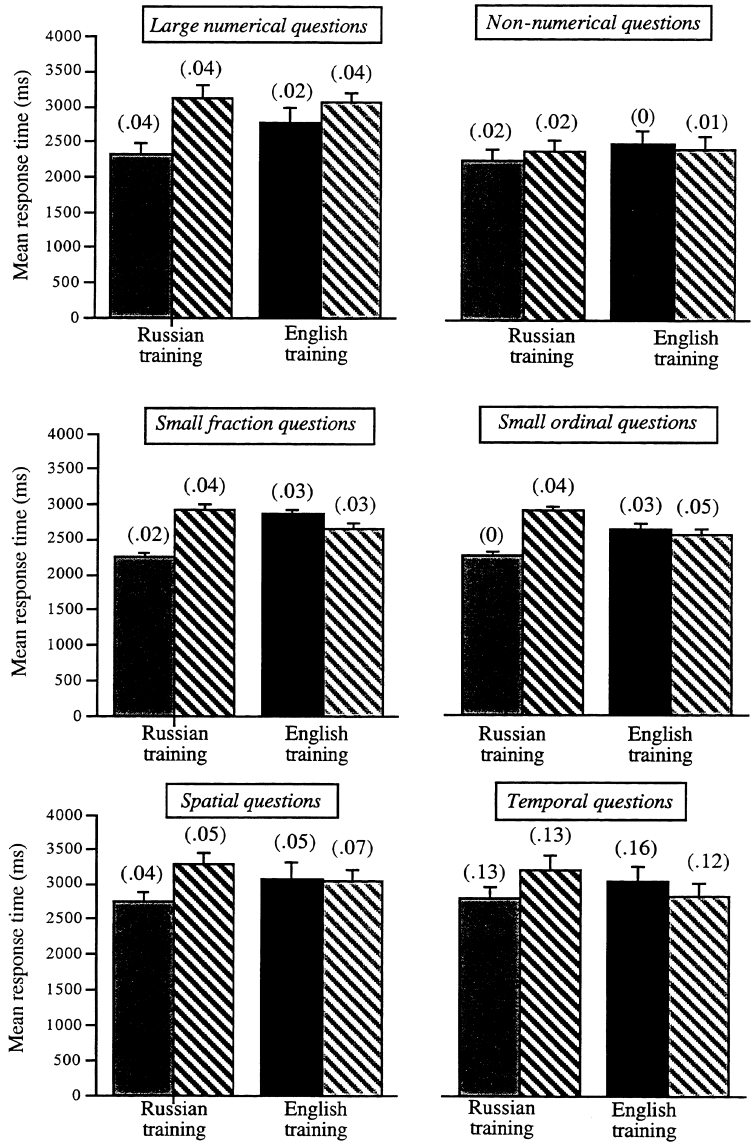

Fig. 8. Response latencies and error rates on each type of problem and in each language in Experiment 3 during testing in the trained and untrained languages (solid and stripped bars, respectively).

whether the training language was Russian or English. For the non-numerical facts, subjects answered with equal speed in the language of training and in the untrained language. Accuracy was high in all conditions.

These effects were confirmed by a 2 (fact type: numerical versus non-numerical) × 2 (language of training: Russian versus English) × 2 (problem novelty:

question asked during training versus not asked) × 2 (language change: same versus different) ANOVA on response latencies. The analysis revealed main effects of fact type ($F(1,7) = 236.82$, $P < 0.001$) and language change ($F(1,7) = 154.84$, $P < 0.001$), and a significant interaction between these factors ($F(1,7) = 17.46$, $P < 0.005$). The only other significant effects in the analysis were an interaction of language of training and language change ($F(1,7) = 9.82$, $P < 0.02$), and an interaction of those two factors with task ($F(1,7) = 5.90$, $P < 0.05$). The latter interactions reflect subjects' tendency to respond faster when tested in Russian, especially after training in Russian and especially for numerical questions.

Next we consider subjects' memory for material involving small numbers. Fig. 8 presents the mean response latencies and error rates for the set of questions testing for information about simple fractions and small ordinal numbers (not included in the category of numerical questions analyzed above). Both sets of questions yielded the same two findings. First, subjects performed faster in Russian than in English. Second, in each test language (Russian or English), subjects performed faster and somewhat more accurately if they were trained on the problems in that language than if they were trained on the problems in their other language. Because the superiority of performance in Russian was slightly greater than the superiority of performance in the trained language, performance on material learned in English was slightly faster during testing in Russian than during testing in English. Accuracy was as high or higher during testing in the training language, indicating no speed–accuracy trade-off. A 2 (fact type: fractions versus ordinals) × 2 (language of training) × 2 (question novelty) × 2 (language change) ANOVA on the response times revealed only a main effect of language change ($F(1,7) = 12.28$, $P < 0.01$), and an interaction of language change by language of training ($F(1,7) = 13.30$, $P < 0.01$), reflecting the above two effects.

Finally, Fig. 8 presents subjects' performance on questions testing their memory for spatial and temporal information within the lessons. Both categories of questions showed the same two effects as did the small number questions: faster responding in Russian than in English and faster responding, in each test language, after training in the same language. Again, the benefit from testing in Russian was slightly greater than the benefit from testing in the trained language, and so subjects performed slightly better on the material learned in English when tested in Russian than when tested in English. For the spatial questions, the language-of-training advantage is not attributable to a speed–accuracy trade-off, because accuracy was as high or higher for problems in the trained language. For the temporal questions, accuracy was slightly higher for problems in the untrained language, complicating the interpretation of the latency differences. The 2 (fact type: spatial versus temporal) × 2 (language of training) × 2 (language change) ANOVA revealed a significant main effect of language change ($F(1,7) = 14.45$, $P < 0.01$), complicated by interactions of language change with fact type ($F(1,7) = 5.81$, $P < 0.05$) and with language of training ($F(1,7) = 6.73$, $P < 0.05$). The effect of language change was greater for the spatial facts and after training in Russian.

4.3. Discussion

Bilingual subjects who learned new numerical facts in the context of a fictitious history or geography lesson subsequently retrieved those facts with greater speed and accuracy in the language of training than in the untrained language. In contrast, non-numerical facts about categories of objects, learned during the same lessons, were subsequently retrieved with equal ease in the two languages.

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 whenever they are learned.

Although representations of large numbers were found to depend on language, subjects nevertheless managed to use these representations to answer questions in their untrained language. In contrast to Experiments 1 and 2, this feat could not be accomplished in Experiment 3 without using information provided during the training sessions in the other language. The fact that subjects did retrieve the correct answers with quite high accuracy in the untrained language provides evidence that transfer of training occurred across languages. It is possible that subjects answered questions in the untrained language by retrieving the information in the trained language and then translating between their two languages. Alternatively, subjects may have formed both language-dependent and language-independent representations of the numerical material during the learning phase, using the former information more quickly at test.

Experiment 3 also provided evidence for language-specific representations of small ordinal and fractional numbers. After subjects learned new facts about simple fractions and ordinals involving only numbers below 5, they subsequently retrieved those facts more effectively in the language of training than in the untrained language. These findings were not predicted from the thesis that humans have a language-independent representation of approximate numerosity that operates in accord with Weber's Law, because such a representation should have been sufficient to distinguish exactly between numerosities of 4 and below.

This unpredicted finding might be explained in three different ways. First, subjects indeed may have a language-independent representation of number that accords with Weber's Law and serves to distinguish small numerosities exactly, but they may fail to use it in Experiment 3 because of a feature of the design of that study. In Experiment 3, subjects learned facts about large and small numbers within the same stories, and they were tested on both kinds of number facts within the same block of trials. In Experiments 1 and 2, in contrast, subjects always were presented with exact and approximate number problems in separate, blocked trials. If subjects adopt a single encoding strategy for answering all numerical questions within a single test block, then only a language-dependent, exact number strategy would have been effective in Experiment 3. Second, subjects may have a language-independent representation of exact small cardinal numbers, but they may lack such a representation of fractions and ordinal numbers. Although human infants and animals have been found to form language-independent representations of the cardinal numbers 'one' through 'four', it is not known whether any animal forms language-independent representations of fractional or ordinal numbers. Third, adults who have learned arithmetic may engage language-dependent processing whenever they encounter problems involving exact numbers, regardless of the size of those numbers.

Whatever the causes of subjects' extensive language specificity for numerical information, the existence of language-specific learning may have implications for learning in bilingual classrooms. When students learn real lessons in history, geography, and other subjects, information about exact large numbers, fractions, and ordinals often are interspersed. The present findings suggest that such conditions will favor the development of language-specific representations of number.

Finally, Experiment 3 provides preliminary information concerning bilingual learning of information about space and time. When subjects learned new facts about the time or place at which an event took place, their performance showed two striking features. First, performance was better in their first language – Russian – than in their second language. Second, performance in each language was better after training in that language than after training in their other language. Although subjects answered questions about space and time faster in Russian than in English, they answered faster in English after training in English than after training in Russian. These findings agree with prior suggestions that humans encode information about space and time in representations that depend, in part, on a specific natural language (Hermer-Vazquez et al., 1999; Levinson, 1996).

5. General discussion

Three experiments provided evidence for language specificity in bilingual learning about numbers. Russian–English bilinguals retrieved information about exact numerosity faster when queried in the language in which they acquired the information, both when they were tested for knowledge of arithmetic facts and when they were tested for knowledge in other domains. The language-of-training advantage cannot be attributed to differences in the speed of encoding and decoding processes, for it occurred with bilingually-trained numbers as well as with monolingually-trained numbers. It cannot be attributed to any habitual preference for processing information in one language, because it occurred with training in English (subjects' non-preferred language) as well as Russian. Finally, the language-of-training advantage does not reflect a global effect of the input language on retrieval of all information, because it was not observed when subjects learned and retrieved non-numerical information or information about approximate numerosity. Rather, it appears that information about exact large numerosity is stored in a language-specific form.

Although Russian–English bilinguals retrieved new arithmetic facts involving approximate numerosities (in Experiments 1 and 2) and new facts involving information about categories of objects and events (Experiment 3) with equal speed and accuracy in both their languages, Experiment 3 suggests that language-specific learning may occur quite extensively. Preliminary findings suggest that subjects formed language-specific representations of space and time as well as number, consistent with suggestions that uniquely human spatial and temporal concepts are language-dependent (Hermer & Spelke, 1996; Hermer-Vazquez et al., 1999; Levinson, 1996; Peacocke, 1992). Moreover, the participants in Experiment 3 showed a language-of-training advantage in answering questions about simple fractions and ordinal numbers. In Experiment 3, subjects who learned fictitious history and geography lessons therefore appeared to encode or retrieve a considerable amount of information in a language-dependent form.

In closing, we discuss one practical question and one theoretical question raised by this research. First, what might be the implications of the present findings for education, and especially for debates concerning the merits of bilingual education for children? Second, what might be the role of language in human representations of exact number?

5.1. Education

In many parts of the world, school classes include students from diverse language backgrounds, and teachers face the task of preparing those students to function in a non-native language environment. Are such children best served by instruction in the language in which they will eventually function, by instruction in their native language, or by instruction in both languages (a bilingual classroom)?

Although we can offer no conclusive answer to this question, we believe that attempts to evaluate different educational strategies in multilingual environments should consider three possibilities suggested by the present research. First, a specific, natural language may serve not only as a medium of input for learning but as a medium of representation for learned information. Second, non-native speakers of the classroom language may learn certain material presented in that language as easily and effectively as native speakers. Third, the costs and benefits of learning in a non-native language may not be uniform across the spectrum of things to be learned. We consider each suggestion in turn.

The present research provides evidence that a specific natural language serves, in part, as a medium of representation for new information about large, exact numbers. If this conclusion applies to children in classrooms, it has some unsettling implications. First, a child who is told that ‘two plus two equals four’ and that ‘deux et deux font quatre’ may need to learn two facts, not one. Children taught arithmetic in bilingual classrooms therefore may face a larger learning task than those in monolingual classrooms, not only in subjects such as reading but also in mathematics. Moreover, a child who learns arithmetic in a monolingual, native language classroom may be at a disadvantage if she is later transferred to a class taught in her second language or if her arithmetic knowledge is tested in the second language, even if she shows facility at the second language in other respects. Children whose arithmetic instruction switches from one language to another literally may lack the representations that allow direct solution of arithmetic problems in the second language. To function in the new language environment, they may need to learn arithmetic anew or engage in time-consuming and potentially confusing translations between their languages.

On the positive side, the present research suggests that some learning may be as effective in a child's poorly mastered, second language as in her first language. The subjects in Experiments 1 and 2 were native speakers of Russian, had learned arithmetic entirely in Russian, and strongly preferred to perform arithmetic calculations in Russian. Nevertheless, they learned new arithmetic facts as quickly and effectively in English as in Russian. This finding raises questions about the prevalent view that children learn better if taught in their native language. At least in the case of arithmetic, children may not need great facility with a language in order to use that language as a medium of representation.

The third suggestion concerning bilingual instruction is perhaps the most important: the relative merits of monolingual versus bilingual instruction may vary depending on the nature of the material to be learned. In particular, a child in a bilingual classroom may be at a disadvantage in acquiring information concerning large, exact numerosities but at no disadvantage in acquiring information about approximate numerosity. Moreover, a child in a non-native language classroom may be equally as able as native-speaking children to learn new facts of arithmetic, but less able to learn new facts about spatial relations, temporal relations, or small numbers. In the present experiments, bilingual subjects showed a language-of-training advantage for some of the facts they learned: exact addition facts and historical and geographical facts involving large numbers. For other materials, subjects showed a first-language advantage. They learned new multiplication facts more effectively in Russian in Experiment 2, and they learned facts about ordinal numbers, fractions, and spatial and temporal relationships more effectively in Russian in Experiment 3. For still other materials, subjects proved equally capable of learning in either language and showed full transfer from one language to the other. These findings suggest that children may acquire a mix of language-dependent and language-independent information in classroom settings, complicating the evaluation of monolingual versus bilingual education.

A final suggestion arising from the present findings applies to children learning mathematics in monolingual as well as bilingual classrooms. Our findings suggest that children may derive considerable benefit from attempts to cultivate their 'number sense' (Dehaene, 1997) by teaching them arithmetic facts about approximate numerosities. Although some number sense is present in animals and human infants, it is clear from our studies that number sense can be enhanced. In two short training sessions, the subjects in Experiments 1 and 2 learned new facts about approximate numerosity. Moreover, learning approximate number facts generalized to new problems involving untrained numbers, whereas learning exact number facts did not. Efforts to develop children's number sense therefore may yield rapid and quite general benefits.

All these suggestions are tentative, for they depend on the untested assumption that learning by children in classroom settings depends on the same representations and processes as learning by adults in laboratory settings. Given the increasing prevalence of multilingual cultures and learning environments and the increasing concern over mathematics education, however, research investigating these suggestions could be of considerable importance.

5.2. Language and thought

We close by considering why representations of exact large numerosity might depend on a specific natural language. At first glance, this finding seems highly counterintuitive. Human knowledge of number appears to be quintessentially abstract. The concept 'seven' appears to transcend any of the particular sets of seven entities that a person enumerates, the particular situations in which she enumerates them, and (one would think) the particular language in which she expresses this enumeration. However, our findings suggest that 'seven' is a language-dependent concept, distinct from the Russian 'sem', or the French 'sept'. Why might concepts like 'seven' depend on a specific natural language?

Hints of an answer to this question come from research in a different domain, on human and non-human representations of the positions of objects in the spatial layout. Research by Cheng (1986) and Gallistel (1990) provides evidence that rats represent the geometric structure of the environment, as well as non-geometric properties of the environment such as odors, surface patterns, and surface brightness. Nevertheless, these two types of representation are not readily combined. Rats fail to conjoin information about space and objects so as to represent that one object bears a particular geometric relation to another. Research by Hermer and Spelke (1994, 1996) provides evidence for similarly modular representations in very young children. In contrast to these findings, however, children begin to show evidence, in their spatial behavior, of an ability to conjoin geometric and non-geometric information at the age at which they begin to produce expressions such as 'The toy is left of the truck' (Hermer-Vazquez, 1997). Most strikingly, human adults who normally show this ability appear temporarily to lose it and to form only separate geometric and non-geometric representations like young children and rats when they engage in a simultaneous verbal interference task (Hermer-Vazquez et al., 1999). The distinctive spatial memory abilities of human adults and older children appear to depend in part on language.

What is the nature of this dependence? Hermer-Vazquez (1997) and Spelke and Tsivkin (in press) have suggested that language serves as a medium for conjoining information from the separate, modular representations that humans share with other animals. Specific natural languages serve this function because they have two properties not found in language-independent, modular representations. First, a natural language is a domain-general medium of representation. It allows speakers to represent information about space, time, objects, colors, odors, people – to a first approximation, any information that their domain-specific cognitive systems make explicit. Second, a natural language is a compositional system. Once speakers have learned a finite set of terms and rules of combination, they can express indefinitely many new expressions with no further learning. In particular, children may learn expressions such as 'The door is left of the wall' by relating those expressions to the purely geometric representations that preverbal children share with rats, and they may learn expressions such as 'the doll' and 'the blue truck' by relating those expressions to language-independent representations of objects. The compositional principles of their language would then allow them to form and understand expressions such as 'The doll is left of the blue truck', with no further learning. If no language-independent system allows such combinations (as suggested by the research of Cheng with rats, and by Hermer and colleagues with preverbal children and with adults experiencing verbal interference), then such representations will be unique to humans and will depend on the acquisition and use of a specific language.

This hypothesis could account for the language-of-training advantage for learning about spatial relationships in Experiment 3, but can it explain why our bilingual subjects formed language-dependent representations of number? Research with animals and preverbal infants provides evidence for language-independent representations of approximate numerosity (Gallistel, 1990), as did Experiments 1 and 2 in the present series. These representations are limited in precision, however, and cannot capture the effects of adding just one object to a sufficiently large array. Research with animals and preverbal infants also provides evidence for language-independent representations of exact numerosities for sets with four or fewer members. For example, monkeys and infants form representations of discrete objects that capture the distinction between one object and two objects and allow computation of the effects of adding exactly one object to an array of objects (Hauser et al., 1996; Wynn, 1992a; for discussion see Hauser & Carey, 1998), although some of the properties of this system are currently under debate (see Scholl, in press).

Although small, exact numerosities could in principle be distinguished by the large approximate number system (see Dehaene, 1997; Gallistel & Gelman, 1992), studies of monkeys and human infants suggest that these representations are constructed by two distinct systems (Carey & Spelke, in press; Hauser & Carey, 1998). First, infants and monkeys are able to perform additions on small numbers of items that are occluded, but they fail to perform additions on large numbers of items when the correct and incorrect numerosities differ by the same ratio. They can add $1 + 1$ to yield 2 rather than 1 but fail to add $5 + 5$ to yield 10 rather than 5 (Feigenson & Carey, 2000; Hauser et al., 2000). This finding and others (see Carey & Spelke, in press) suggest that the large approximate system fails to represent each member of a set as a persisting individual. Second, research provides evidence that infants represent the invariant cardinal values of large sets over changes in continuous variables such as the sizes and spacing of individual elements (Xu & Spelke, 2000) but fail to represent the cardinal values of small sets over the same changes (Clearfield & Mix, 1999; Feigenson, Carey, & Spelke, 1998; Xu, 2000). This finding and others (see Carey & Spelke, in press) suggest that the small number system fails to represent a group of individuals explicitly as a set. If infants and non-human primates are unable spontaneously to combine their small and large number systems, therefore, they will be able to represent specific individuals and specific sets of non-individuated entities, but they will be unable to represent specific sets of individuals.

We suggest that a natural language counting system allows humans to combine these two distinct types of representation into a single, language-dependent representation of discrete, exact numerosity: a representation of sets of individuals whose cardinality increases as new individuals are added to the set. The new, hybrid system captures the benefits of both initial, language-independent systems and overcomes the specific limits of each system. Like the exact small number system, it is discrete, exact, and applies to persisting individuals; like the approximate large number system, it serves to represent sets with no upper bound and with explicit cardinal values. Because language serves to link the two types of representations that compose this system, however, representations within this system depend on it.

Research on the development of understanding of number words and of verbal counting suggests how the young child's initial representations may combine to form distinctively human representations of number (Wynn, 1990, 1992b). When children first begin to produce the number words, they map these words only onto the distinction between singular and plural: 'one' refers to *an individual*, and 'two', 'three', 'six', and all other number words refer indiscriminately to *a set larger than one*. At this stage, 'one' may connect only to representations of individuals constructed within the exact small number system, and the other number words may connect only to representations of sets constructed by the approximate large number system. As children experience the different number words in the same quantifier positions in sentences (Bloom & Wynn, 1997), they may come to connect each number word to representations in both systems: representations of *a set of individuals*. Moreover, as they use the number words in sequence in the counting routine, they may come to appreciate that each word in the counting sequence picks out a *set with one more individual* than the previous word. These developments would complete the child's construction of the natural numbers.

The hypothesis that natural number concepts are constructed by combining two numerical representations and that natural language serves as the medium for this combination would account for the present findings, but other accounts also are possible. In particular, children may use number words to express number concepts that are antecendently available (Bloom & Wynn, 1997), and adults may use number words in arithmetic and other memory tasks because those words increase the accessibility or efficiency of those number representations. As a further possibility, children may use language to construct a system of number representation by combining a single language-independent system with a set of concepts provided by the language faculty itself: the concepts that provide the semantics of natural language quantification (Carey & Spelke, in press; Hurford, 1987). Further research probing in greater depth how children learn number words, and how adults use them, is needed to distinguish these possibilities.

All of the above views predict that small, exact numbers and large, approximate numbers can be represented independently of language, and that only representations of exact large numerosities depend on a specific language with a counting system. Part of this prediction was supported by Experiments 1 and 2. Learning new facts about exact large numbers was found to be language-dependent, whereas learning new facts about approximate large numbers was not. Experiment 3 partially confirmed this prediction as well, because subjects showed a language-of-training advantage for numerical information and spatial information but not for information about object categories. Two findings from Experiment 3, however, failed to agree with this prediction. Subjects showed a language-of-training advantage when they were asked about small, simple fractions and about ordinal numbers below 5. Because the small numbers were presented as fractions or ordinals, it remains an open question as to whether information about small cardinal numbers can be learned and retained in a language-independent manner.

The present experiments suggest a general approach to questions about the sources of uniquely human cognitive abilities. Compared to other mammals, humans have very similar perceptual and action systems and very similar systems for getting around in space, orienting in time, recognizing objects, and negotiating social encounters. Nevertheless, humans' cognitive achievements far outpace those of any other animal, especially in formal domains like mathematics. The flexibility of human cognition is particularly striking. Although all animals look intelligent when they solve problems for which their cognitive systems have evolved, humans often act intelligently when they confront new problems of their own design. Such intelligence, we suggest, stems in part from a strategy that humans use again and again: we combine old concepts and procedures together to form new ones. Natural language, our most striking combinatorial system, may provide one of the tools that makes this strategy possible. Formal mathematics may be one of its richest and most dramatic outcomes.

Acknowledgements

Portions of this research were submitted as a masters thesis at MIT by S.T.; portions of Experiments 1 and 2 were reported in *Science* (Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999). We thank Stanislas Dehaene, Susan Carey, Frank Keil, Molly Potter, Karen Wynn, and Fei Xu for criticisms and suggestions and Kirsten Condry for assistance. This study was supported by NIH grant 23103 to E.S.S.

Appendix A

A.1. Experiment 1: example problems with bilingually-trained responses

Add 54:
54 + 48 = 102 versus 92; 54 + 76 = 130 versus 140; 54 + 88 = 142 versus 152
Add 63:
63 + 39 = 102 versus 92; 63 + 67 = 130 versus 140; 63 + 79 = 142 versus 152
Add in base 6:
2 + 5 = 11 versus 21; 5 + 15 = 24 versus 34; 13 + 15 = 32 versus 33
Add in base 8:
4 + 5 = 11 versus 21; 6 + 16 = 24 versus 34; 27 + 3 = 32 versus 33
Cube root estimation:

cube root of 29 $\approx$ 3 versus 4; cube root of 65 $\approx$ 4 versus 3; cube root of 120 $\approx$ 5 versus 7

Base 2 log estimation:

log base 2 of 9 $\approx$ 3 versus 4; log base 2 of 15 $\approx$ 4 versus 3; log base 2 of 33 $\approx$ 5 versus 7

A.2. Experiment 2: example problems with bilingually-trained responses

Exact addition:

$34 + 71 = 105$ versus 115; $86 + 26 = 112$ versus 102; $49 + 79 = 128$ versus 118

Approximate addition:

$34 + 71 \approx 110$ versus 80; $86 + 26 \approx 120$ versus 90; $49 + 79 \approx 120$ versus 150

Exact multiplication:

$21 \times 5 = 105$ versus 115; $28 \times 4 = 112$ versus 102; $16 \times 8 = 128$ versus 118

Approximate multiplication:

$21 \times 5 \approx 110$ versus 80; $28 \times 4 \approx 120$ versus 90; $16 \times 8 \approx 120$ versus 150

A.3. Experiment 3: example problems

Exact large numbers, history:

There were ___ elected governors in Kapnopa prior to the execution. (57 versus 58)

Governor Pelba predicted that it would be ___ days before the farmers would be forced out of business. (20 versus 22)

How many times did the acrobat flip? (12 versus 10)

Exact large numbers, geography:

On the stone, Mary discovered ___ lines of Peaken text. (57 versus 58)

When Mary started looking for the treasure, she was ___ years older than when she first encountered the Iuto River. (20 versus 22)

How many scattered reports of the discovery of treasure did Mary hear about? (12 versus 10)

Object categories, history:

The fruit vendors bought ___ and grapes from the Kapnopan farmers. (pears versus strawberries)

Where did a band of farmers meet? (in a cave versus in a house)

The King's palace was encrusted with ___. (leaves versus flowers)

Object categories, geography:

On her journey, Mary encountered squirrels scurrying in wilted fields of ___. (pears versus strawberries)

Where did Mary find herself after collapsing suddenly? (in a cave versus in a house)

After Mary resolved to hike to the Dossi River, she burst through the ___. (leaves versus flowers)

Exact small numbers, history:

What part of the passage was manned by the King's soldiers? (2/3 versus 3/4)

After which election in Kapnopa did the King execute some of the governors? (first versus second)

The cluster of pearls featured on the crown was ___ of the King's collection of pearls. (1/3 versus 1/4)

Exact small numbers, geography:

What part of all recorded ancient legends of the river did Mary document herself? (2/3 versus 3/4)

Mary was swept away by the Iuto River when she approached it for the ___ time. (first versus second)

The Pex welcomed ___ of the country's migrant birds. (1/3 versus 1/4)

Spatial directions, history:

On which side of the central palace were the guards at the watchtower? (to the right versus to the left)

In which direction across Juja did the governors flee? (north versus south)

Where was the crown's emerald in relation to the cluster of pearls? (above them versus below them)

Spatial directions, geography:

After Mary thought she understood the legend of the nymphs, she proceeded to her ___ upstream the Iuto. (right versus left)

In which direction in relation to the mountains did the Rivers Iuto and Dossi extend? (north versus south)

Where was the map supposed to appear in relation to the nymphs? (above them versus below them)

Time, history:

In which season was Independence Day in Kapnopa celebrated? (summer versus spring)

Pelba and the farmers arrived at a plan by ___. (early morning versus late morning)

When did the execution of Kapnopan governors take place? (at night versus during the day)

Time, geography:

When did the Pex welcome the country's migrant birds? (in the summer versus in the spring)

Mary began looking for the treasure ___. (early in the morning versus in the late morning)

When did Mary climb through the mountains of Nedu? (at night versus during the day)

References

Bloom, P. (1994). Generativity within language and other cognitive domains. *Cognition*, *51* (2), 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. (1993). *The development of numerical competence: animal and human models*. Hillsdale, NJ: Erlbaum.

Brannon, E. M., & Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science*, *282* (5389), 746–749.

Carey, S., & Spelke, E. S. (1994). Domain-specific knowledge and conceptual change. In L. Hirschfeld, & S. Gelman (Eds.), *Mapping the mind: domain specificity in cognition and culture* (pp. 169–200). Cambridge, MA: Cambridge University Press.

Carey, S., & Spelke, E. S. (in press). Constructing the integer list representation of number. In L. Bonati, J. Mehler, & S. Carey (Eds.), *Developmental cognitive science*. Cambridge, MA: MIT Press.

Cheng, K. (1986). A purely geometric module in the rat's spatial representation. *Cognition*, *23* (2), 149–178.

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

Church, R. M., & Meck, W. H. (1984). The numerical attribute of stimuli. In H. L. Roitblat, T. G. Bever, & H. S. Terrace (Eds.), *Animal cognition.* Hillsdale, NJ: Erlbaum.

Clearfield, M. W., & Mix, K. S. (1999). Number versus contour length in infants' discrimination of small visual sets. *Psychological Science*, *10*, 408–411.

Davis, H., & Pérusse, R. (1988). Numerical competence in animals: definitional issues, current evidence, and a new research agenda. *Behavioral and Brain Sciences*, *11* (4), 561–615.

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

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

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science*, *284*, 970–974.

Ellis, N. C., & Hennelly, R. A. (1980). A bilingual word length effect: implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology*, *71*, 43–52.

Feigenson, L., & Carey, S. (2000, July). *Spontaneous ordinal judgments in preverbal infants*. Poster presented at the biennial meeting of the International Conference on Infant Studies, Brighton, UK.

Feigenson, L., Carey, S., & Spelke, E. S. (1998, April). *Numerical knowledge in infancy: the number/mass distinction*. Poster presented at the biennial meeting of the International Conference on Infant Studies, Atlanta, GA.

French-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: MIT Press.

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

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Hove: Lawrence Erlbaum Associates.

Gelman, R. (1991). Epigenetic foundations of knowledge structures: initial and transcendent constructions. In S. Carey, & R. Gelman (Eds.), *The epigenesis of mind: essays on biology and cognition* (pp. 293–322). Hillsdale, NJ: Erlbaum.

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

Gonzalez, E. G., & Kolers, P. A. (1987). Notational constraints on mental operations. In G. Deloche, & X. Seron, *Mathematical disabilities: a cognitive neuropsychological perspective* (pp. 27–42). Hillsdale, NJ: Erlbaum.

Hauser, M. D., & Carey, S. (1998). Building a cognitive creature from a set of primitives: evolutionary and developmental insights. In C. Allen, & J. Cummins (Eds.), *The evolution of mind.* Oxford: Oxford University Press.

Hauser, M. D., Carey, S., & Hauser, L. (2000). Spontaneous number representation in semi-free-ranging rhesus monkeys. *Proceedings of the Royal Society*, *267*, 829–833.

Hauser, M. D., MacNeilage, P., & Ware, M. (1996). Numerical representations in primates. *Proceedings of the National Academy of Sciences USA*, *93*, 1514–1517.

Hermer, L., & Spelke, E. S. (1994). A geometric process for spatial reorientation in young children. *Nature*, *370* (6484), 57–59.

Hermer, L., & Spelke, E. S. (1996). Modularity and development: the case of spatial reorientation. *Cognition*, *61*, 195–232.

Hermer-Vazquez, L. (1997). *Cognitive flexibility as it emerges over evolution and development: the case of human spatial reorientation*. Unpublished doctoral dissertation, Cornell University.

Hermer-Vazquez, L., Spelke, E. S., & Katsnelson, A. S. (1999). Sources of flexibility in human cognition: dual-task studies of space and language. *Cognitive Psychology*, *39*, 3–36.

Holender, D., & Peereman, R. (1987). Differential processing of phonographic and logographic single-digit numbers by the two hemispheres. In G. Deloche, & X. Seron (Eds.), *Mathematical disabilities: a cognitive neuropsychological perspective* (pp. 43–85). Hillsdale, NJ: Erlbaum.

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

Kolers, P. (1968). Bilingualism and information processing. *Scientific American*, *218*, 78–86.

Levinson, S. (1996). Frames of reference and Molyneux's question: cross-linguistic evidence. In P. Bloom, M. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space*. Cambridge, MA: MIT Press.

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* (6014), 57–59.

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.

Peacocke, C. (1992). *A study of concepts*. Cambridge, MA: MIT Press.

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

Putnam, H. (1980). What is innate and why. In M. Piatelli-Palmarini (Ed.), *Language and learning: the debate between Piaget and Chomsky*. Cambridge, MA: Harvard University Press.

Scholl, B. (Ed.) (in press). Special issue on objects and attention. *Cognition*.

Spelke, E. S., & Tsivkin, S. (in press). Initial knowledge and conceptual change: space and number. In M. Bowerman, & S. Levinson (Eds.), *Language acquisition and conceptual development*. Cambridge, MA: Cambridge University Press.

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. (1992a). Addition and subtraction by human infants. *Nature*, *358* (6389), 749–750.

Wynn, K. (1992b). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, *24* (2), 220–251.

Xu, F. (2000, July). *Numerical knowledge in infancy: two systems of representation*. Paper presented at the biennial meeting of the International Conference on Infant Studies, Brighton, UK.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*, B1–B11.