

## Empirical Research

**Natural Number Bias in Arithmetic Operations With Missing Numbers – A Reaction Time Study**Konstantinos P. Christou<sup>\*a</sup>, Courtney Pollack<sup>b</sup>, Jo Van Hoof<sup>c</sup>, Wim Van Dooren<sup>c</sup>

[a] Early Education Department, University of Western Macedonia, Florina, Greece. [b] Massachusetts Institute of Technology, Cambridge, MA, USA. [c] Centre for Instructional Psychology and Technology, KU Leuven, Leuven, Belgium.

**Abstract**

When reasoning about numbers, students are susceptible to a natural number bias (NNB): When reasoning about non-natural numbers they use properties of natural numbers that do not apply. The present study examined the NNB when students are asked to evaluate the validity of algebraic equations involving multiplication and division, with an unknown, a given operand, and a given result; numbers were either small or large natural numbers, or decimal numbers (e.g.,  $3 \times \_ = 12$ ,  $6 \times \_ = 498$ ,  $6.1 \times \_ = 17.2$ ). Equations varied on number congruency (unknown operands were either natural or rational numbers), and operation congruency (operations were either consistent – e.g., a product is larger than its operand – or inconsistent with natural number arithmetic). In a response-time paradigm, 77 adults viewed equations and determined whether a number could be found that would make the equation true. The results showed that the NNB affects evaluations in two main ways: a) the tendency to think that missing numbers are natural numbers; and b) the tendency to associate each operation with specific size of result, i.e., that multiplication makes bigger and division makes smaller. The effect was larger for items with small numbers, which is likely because these number combinations appear in the multiplication table, which is automatized through primary education. This suggests that students may count on the strategy of direct fact retrieval from memory when possible. Overall the findings suggest that the NNB led to decreased student performance on problems requiring rational number reasoning.

*Keywords:* natural number bias, rational numbers, multiplication makes bigger, misconception, operation

Journal of Numerical Cognition, 2020, Vol. 6(1), 22–49, <https://doi.org/10.5964/jnc.v6i1.228>

Received: 2019-05-23. Accepted: 2019-12-23. Published (VoR): 2020-06-15.

\*Corresponding author at: 3rd km. Florina – Niki, 53100, Florina, Greece. E-mail: [kchristou@uowm.gr](mailto:kchristou@uowm.gr)



This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License, CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rational number knowledge is an important component of mathematical learning and predicts later mathematical achievement (Bailey, Hoard, Nugent, & Geary, 2012; Siegler, Fazio, Bailey, & Zhou, 2013). Yet, across age and nationality, individuals struggle with understanding and using rational numbers in different mathematical contexts (e.g. Behr, Harel, Post, & Lesh, 1994; Gómez, Jiménez, Bobadilla, Reyes, & Dartnell, 2014; Iuculano & Butterworth, 2011; Mazzocco & Devlin, 2008; Vamvakoussi, Christou, Mertens, & Van Dooren, 2011). Difficulties may appear because prior knowledge and experience with natural numbers often does not support rational number learning. Natural numbers are based on different principles and properties than rational numbers, such that application of natural number properties and rules when reasoning with rational numbers may lead to misconceptions and errors (Carpenter, Fennema, & Romberg, 1993; Ni & Zhou, 2005; Smith, Solomon, & Carey, 2005; Vamvakoussi & Vosniadou, 2010). The current study investigates how prior natural number knowledge may result in a well-documented misconception in the mathematics education community: The tendency for students to associate arithmetic operations with specific result sizes, i.e., bigger numbers in

multiplication and smaller numbers in division (Fischbein, Deri, Nello, & Marino, 1985; Greer, 1987; Izsák & Beckmann, 2018; Onslow, 1990; Prediger, 2008).

## The Natural Number Bias Phenomenon

The fact that prior number knowledge interferes with learning more advanced number concepts has been well-established in mathematics education for decades (e.g. Hart, 1981; Rees & Barr, 1984). From a developmental perspective, this phenomenon is called the *whole number bias* (Ni & Zhou, 2005) or more recently the *natural number bias* (NNB) (Vamvakoussi, Van Dooren, & Verschaffel, 2012; Van Dooren, Lehtinen, & Verschaffel, 2015), and refers to students' tendency to use natural number knowledge when reasoning about rational numbers. The NNB can explain student difficulties and erroneous behaviors in different number task domains. For example, when reasoning about rational number magnitudes, students may erroneously think that – as with natural numbers – decimal numbers with more digits are larger, e.g., that 2.367 is larger than 2.6 (Moutier & Houdé, 2003; Nesher & Peled, 1986; Resnick et al., 1989; Roell, Viarouge, Houdé, & Borst, 2017, 2019). Similarly, students tend to think that the bigger the numerator and denominator of a fraction, the bigger the fraction value, which results in mistakes in ordering fractions, e.g., 249/1000 is larger than 1/4 (DeWolf & Vosniadou, 2011; Hartnett & Gelman, 1998; Moss, 2005). As another example, the NNB may create difficulties for students to understand that the set of rational numbers is dense (i.e., that there are infinitely many numbers between any two rational numbers). Because the set of natural numbers is discrete (i.e., between two successive natural numbers there is no other natural number), the NNB may lead students to think that between two pseudo-consecutive numbers (e.g., 0.5 and 0.6), there is no other number (Desmet, Grégoire, & Mussolin, 2010; Vamvakoussi & Vosniadou, 2010).

Most often associated with the NNB is students' reasoning about the size of the result of rational number arithmetic. Students tend to think that multiplication always results in a larger number while division always results in a smaller number. Well-known to teachers, this misconception has been reported in the literature since the 1920s (Thorndike, 1922 as mentioned by Krueger & Hallford, 1984). Empirical research has shown that this misconception systematically appears when students solve word problems (Bell, Swan, & Taylor, 1981; Dixon, Deets, & Bangert, 2001; Fischbein et al., 1985; Graeber, Tirosh, & Glover, 1989; Greer, 1987, 1994; Harel & Confrey, 1994; Hart, 1981) and occurs across schooling. As examples, a majority of second graders have incorrectly answered whether  $4.6 \div 0.6$  is more or less than 4.6 (Greer, 1987), secondary students have responded that  $x > x \times 2$  cannot be true (Van Hoof, Vandewalle, Verschaffel, & Van Dooren, 2015), and college students have responded that  $z \times 7$  cannot be smaller than 7 (Vamvakoussi, Van Dooren, & Verschaffel, 2013).

Fischbein et al. (1985), who offered one of the first insights for this phenomenon, suggested that students hold implicit intuitive models of arithmetic, which shape their expectations about the effect of operations. These intuitive models associate addition with putting together, subtraction with taking away, multiplication with repeated addition, and division with equal sharing. From a NNB perspective, Vamvakoussi and colleagues (2013) argued that these intuitive models are compatible with - and based on - natural number operations, and relate to the operation and not the numbers involved. Specifically, addition and multiplication between natural numbers (excluding 0 and 1) always results in larger numbers. Similarly, the result of subtraction or division between two natural numbers is always a smaller number. This is, however, not true for non-natural numbers, for which the effects of operations depend on the numbers involved. Multiplication with rational numbers less

than 1 results in answers that are smaller than at least one operand (e.g.,  $4 \times 0.2 = 0.8$ ) and division with rational numbers less than 1 results in answers that are larger than the operands (e.g.,  $8 \div 0.4 = 20$ ).

Students' years of experience working with natural numbers may create a strong intuition about the results that can be expected from natural-number arithmetic (e.g., *multiplication makes larger*) (Greer, 1994), or *addition makes larger* (Dixon et al., 2001). From a framework theory approach to conceptual change, students from very early on construct an initial understanding for the number concept, which, based on their experience with counting and the sequence of the number words, acquires the characteristics of the mathematical concept of natural numbers (Gelman, 2000; Smith et al., 2005; Vamvakoussi, Christou, & Vosniadou, 2018; Vosniadou, Vamvakoussi, & Skopeliti, 2008). This initial number concept, which is organized into a framework theory for number, may form students' beliefs, their interpretations and their anticipations about the properties of numbers. From this perspective, the influence of the initial number concept when reasoning with more advanced numbers such as the rational and the real numbers may result in the NNB phenomenon and lead to systematic errors in rational number tasks where the properties of natural numbers do not hold (Vamvakoussi et al., 2018).

Importantly, the initial number conception remains even after learners have acquired the correct mathematical knowledge about rational numbers and, in principle, are able to solve rational number tasks. Thus, natural number knowledge may still interfere after students have learned to reason with rational numbers and have acquired the knowledge to solve rational number tasks (see also Shtulman & Valcarcel, 2012; Vosniadou et al., 2008). Support for this idea, and specifically for the NNB with arithmetic operations, comes from studies using accuracy and reaction time measurements. If properties of natural numbers *come to mind first*, even in learners who have the correct rational number knowledge, responses to arithmetic equations that are aligned with natural number knowledge (called congruent items) would be quicker and more accurate than for incongruent items, in which natural number knowledge leads to a false conclusion. This is because a learner has to overcome this initial idea using analytical reasoning that may not always be successful, which in turn produces longer response times and fewer accurate responses. These response patterns align with dual process theories that rely on the distinction and interaction between intuitive and analytical reasoning processes (Gillard, Van Dooren, Schaeken, & Verschaffel, 2009; Vamvakoussi et al., 2013). Several reaction time studies with student and adult populations support the theories of co-existence and intuitive interference of natural number knowledge by showing statistically significant differences in accuracy rates and response times between tasks that were in-line with intuitions about the results of arithmetic operations and tasks that falsified these intuitions (Obersteiner, Van Hoof, Verschaffel, & Van Dooren, 2016; Vamvakoussi et al., 2012; Van Hoof et al., 2015).

## The Dual Effect of NNB in Arithmetic Operations With Missing Numbers

Students' initial conception of numbers, coupled with their intuitions for the results of arithmetic operations, suggests the NNB may affect how students solve arithmetic problems with missing operands. To test this, Christou administered paper and pencil tasks to primary (Christou, 2015a, 2015b) and secondary school students (Christou, 2017). The tasks included arithmetic equations with operations between a given number and a missing number (e.g.,  $9 \div \_ = 4$ ). Different categories of tasks captured the potential duality of the NNB effect. The three categories included: a) congruent tasks, in which the results of the operations were in-line with students' intuitions (i.e., bigger results for multiplication and smaller for division), and the missing numbers were natural numbers (e.g.,  $3 \times \_ = 12$ ); b) tasks which were again in-line with students' intuitions about the results of

the operations, but the missing numbers were rational numbers larger than 1 instead of natural numbers (e.g.,  $3 \times \_ = 11$ ); and c) tasks in which the magnitude of the results of the operations falsified students' intuitions that multiplication makes bigger and division makes smaller (e.g.,  $3 \times \_ = 2$ ), and in which the missing number was a rational number smaller than 1. In all those studies (Christou, 2015a, 2015b, 2017), students had statistically significantly higher accuracy rates for the items aligned with their intuitions about the size of the results of each operation than items that falsified these intuitions. Additional studies have shown similar performance differences (Obersteiner et al., 2016; Vamvakoussi et al., 2013; Van Hoof et al., 2015).

Importantly, these results also indicated that students tend to think of the missing numbers in the items as natural numbers. Students performed significantly better when the missing number was a natural number than when it was a rational number, even when their intuitions about the results of the operations were not violated (Christou, 2015a, 2015b, 2017). Students' also employ intuitions about missing numbers as natural numbers when evaluating algebraic expressions that contain literal symbols. Christou and colleagues found that students tend to substitute mostly natural numbers for literal symbols when evaluating algebraic expressions with operations between numbers and literal symbols (Christou & Vosniadou, 2012; Christou, Vosniadou, & Vamvakoussi, 2007). For example, students thought that  $k + 3$  represented only natural numbers larger than three. Additionally, in an interview study with tenth graders, the majority of the students claimed that  $5d$  is always bigger than  $4/d$  because multiplication makes the numbers bigger than division, and most students supported this claim by substituting natural numbers for the literal symbols, despite the hints provided by the interviewer to also try with other kinds of numbers (Christou & Vosniadou, 2012). Similarly, Van Hoof and colleagues (2015) found that students who were interviewed, explicitly referred to general rules about the results of the operations that are valid only for natural numbers, or substituted literal symbols with natural numbers to come to an answer.

Strictly speaking, thinking of missing numbers or literal symbols as natural numbers only would not fall under the definition of the NNB as relying on the *properties* of natural numbers when reasoning about rational numbers. However, it would indicate the dominance of natural numbers in students' thinking in numerical situations, and as such could be seen as another instance of the NNB. Thus, these studies suggest a dual effect of the NNB: intuitions about results of arithmetic operations and intuitions that missing numbers or literal symbols represent natural numbers.

## The Present Study

The present study seeks to further examine and disentangle the dual effect of the NNB by measuring participants' accuracy and reaction times as they make evaluations about whether arithmetic equations with a missing operand could be true or not (e.g., Is there a number such that  $3 \div \_ = 12$ ?). In these tasks, the NNB may affect participants' evaluations about the validity of such equations in two main ways. First, it would affect their strategy to check the validity of the equations by mentally substituting specific numbers for the missing number symbols. Under the influence of the NNB, participants may disproportionately substitute natural numbers. Second, participants may make evaluations based on intuitive expectations about the size of the results of operations (e.g., that multiplication should provide a larger outcome than the operand numbers), as shown in prior studies with literal symbols as missing numbers (Obersteiner et al., 2016; Vamvakoussi et al., 2012, 2013; Van Hoof et al., 2015).

The study has three aims:

1. As reviewed above, prior studies have focused on interviews and paper and pencil tasks that measure student accuracy. This study tests the dual effect assumption by measuring a different aspect of student reasoning: timed evaluations about missing number operands in arithmetic equations. This allows us to judge whether the NNB effect is still present, even when participants give correct responses.
2. It disentangles the two aspects of the NNB that may appear in tasks for which finding the missing number is more or less difficult.
3. It seeks to clarify whether participants' tendencies to correctly evaluate the validity of statements like  $3 \times \_ = 12$  relate to a trial-and-error mental process with specific natural numbers, or rather relate to retrieval of multiplication facts from long term memory.

Based on the dual aspect of the NNB in operations with missing numbers, participants may evaluate the validity of equations such as  $3 \div \_ = 12$  based on two main strategies. First, participants who are affected by the NNB may tend to respond that operations with larger results in multiplication and smaller in division would be evaluated as *possible*, and those with larger results in division and smaller in multiplication would be evaluated as *impossible*. Second, some participants may try to find the specific missing number that makes the expression possible in a trial-and-error process. Thus, participants may indicate that there is a missing number that would make the equation true when the missing natural number is easy to find and would incorrectly respond that *there is no number that makes the equation true* when they cannot immediately find a natural number that would make the equation hold, either because the missing number is not a natural number, or because finding it is difficult. For this reason, we included three groups of equations with varying difficulty for finding the missing number:

- a. the Group of Small Numbers, with numbers smaller than one hundred, where the missing number is easy to find when the missing number is a natural number,
- b. the Group of Decimal Numbers, with equations using decimal numbers, and
- c. the Group of Large Numbers, with numbers larger than one hundred.

In the last two groups of items, the missing number is more difficult to find than in the first group, even when in all cases the missing number is a natural number. As we explain below, differences in participants' accuracy and reaction time among the three groups of items would provide further insights about the underlying cognitive processes when reasoning about the results of such arithmetic operations.

First, an important reason to include a set of items involving decimal numbers is to investigate a context effect. If students tend to think of an unknown number as a natural number in the first place, this tendency may be stronger when all numbers that are given in the equation are natural too. Students may be reminded to think of the unknown number as possibly non-natural when the equations contain decimals. Indeed, recent findings from an empirical study which used the same design of tasks in a paper and pencil condition showed that students are affected by the types of numbers (either natural or decimals) that appear in arithmetic operations (Christou, 2017).

Second, the three sets of items are all needed to achieve the third aim: to clarify whether participants' tendency to correctly evaluate the validity of statements like  $3 \times \_ = 12$  may be through trial-and-error or because of familiarity with multiplication facts. To achieve this, this study included arithmetic combinations that would be familiar to participants because specific arithmetic facts come from the multiplication table or because the

arithmetic facts were easily confused with familiar arithmetic combinations (a false familiarity, e.g.,  $3 \div \_ = 12$ ). Krueger (1986) showed that when there is very good verbatim memory for particular multiplication equations, participants count not only on a plausibility evaluation of the given statements, but also direct fact retrieval from memory (see also Stazyk, Ashcraft, & Hamann, 1982). To test for plausibility evaluation and retrieval mechanisms, the current study included arithmetic equations with small number combinations that appear in the multiplication table and thus draw on strong verbatim memory, and equations with decimals and large natural numbers where there is less or no verbatim memory to support specific evaluations. For the group of Decimal and the group of Large numbers, participants may only count on exact calculations in a mental trial-and-error process, or on plausibility evaluations, which are based on *operational patterns* (see Prather & Alibali, 2008) that may determine the expected results from each operation independently of the numbers involved. Considering the first strategy, calculations especially with large numbers are more difficult than with small numbers (Ashcraft, 1992), and the NNB may affect the second strategy.

Finally, as part of the third aim, this study tested the effect of *false familiarity* in participants' evaluations with equations that appeared familiar, but were not. In these cases, a false result may be difficult to reject because the equation is similar to a true equation but with a different operation. For example, in a *cross-operation confusion* condition (Krueger & Hallford, 1984) people find it difficult to reject such equations as  $7 + 1 = 6$  for which the answer would be true if the same pair of numbers would appear under subtraction (Ashcraft & Battaglia, 1978), and  $3 \div 3 = 9$ , for which the answer would be true with multiplication (Winkelman & Schmidt, 1974). Similar research has shown that participants tend to think that  $16 \div 32 = 2$  is correct (Bell et al., 1981). Additionally, Stazyk and colleagues (1982) reported slower response times and increased error rates in confusing problems such as  $7 \times 4 = 21$  or  $7 \times 4 = 35$ , than in non-confusing problems (e.g.,  $7 \times 4 = 18$ ), in which *confusion problems* were characterized as those with answers adjacent or near the correct answer in the multiplication table, that is answers differing  $\pm 1$  in one of the operands. Both confusion effects stemmed from relatedness among the numbers in the memory representation of multiplication facts. To test for the effect of false familiarity, the current study included a category of items in which the missing numbers were unit fractions. These equations falsified participants' intuitions about the size of the results of multiplication and division with the missing number. For example, in the equation  $7 \div \_ = 42$ , the missing number is  $1/6$  and this makes it seem familiar to  $7 \times \_ = 42$  (where the missing number is 6), which appears in the multiplication table.

In sum, the current study examined accuracy and reaction time for evaluations about arithmetic equations with missing operands. Items were in-line with or violated participants' intuitions about the size of the results of the given operations (operation-congruent/incongruent items) and also about whether only natural numbers could be substituted for the missing number symbols (number-congruent/incongruent items). Also, differences in the effect of each aspect of the NNB were tested by comparing participants' performance across three groups of items that were designed with the above characteristics: The Group of Small numbers, the Group of Decimal Numbers, and the Group of Large Numbers. In order to account for the effect of false familiarity, participants' performance was examined with number-incongruent and operation-incongruent items in which the missing number was a unit fraction.

## Method

### Sample

Seventy-seven college participants participated in this experiment. The participants were mostly first year bachelor students of Educational Sciences, or were graduate students in Education or Art. Participants' age ranged from 18 to 29 years ( $M = 19.16$  years), and 64 defined themselves as female. The students participated in the experiment as a prerequisite for taking the exams in one of their main courses.

### Stimuli

There were 108 stimuli, consisting of 72 experimental and 36 distractor items. Each item consisted of a multiplication or division equation with a given number and a missing number (e.g.,  $3 \times \_ = 12$ ). Missing number symbols were used instead of literal symbols (e.g.,  $x$ ) to discourage participants from applying a general insight about the solvability of linear equations, such as the general solution of  $x=b/a$  for any equation of the form  $a \times x = b$ ,  $a \neq 0$ . Using missing number symbols may discourage participants from using this strategy of equation solving. For each item, participants were asked to evaluate whether it is possible for the equation to be true or not, i.e., whether it would be possible to find a missing number that would provide the given result, without the need to actually find the missing number.

### Experimental Items

The experimental items were generated using a 3 (group type: Small, Decimal, Large Numbers)  $\times$  4 (number-congruent/incongruent  $\times$  operation-congruent/incongruent)  $\times$  2 (operation: multiplication or division) design. Examples of each type are presented in [Table 1](#).

### The Three Groups of Items

Items were organized into three Groups, based on whether small numbers, decimal numbers, or large numbers were operators and outcomes. The Group of Small Numbers included items with small natural number combinations (smaller than 100). Participants are more familiar with these numbers since they are used in the classroom and in everyday life. Multiplication and division between numbers smaller than 100 also appear in the multiplication table, which participants have used extensively throughout schooling. The Group of Decimal Numbers consisted of multiplication and division where the operand number and the outcome were decimal numbers. Lastly, the Group of Large Numbers included number combinations greater than those in the multiplication table. Verbatim memory could support participants' evaluations with small numbers, but not with decimals or large numbers.

### The Four Types of Items

Number-congruent items were equations in which a missing natural number would make the equation true. Number-incongruent items were equations in which a missing rational number would make the equations true. The operation-congruent items were aligned with properties of natural number arithmetic: multiplication equations with a larger outcome or division equations with a smaller outcome. Operation-incongruent items were either multiplication equations with a smaller outcome or division equations with a larger outcome. Both types of equations were true.

Table 1

*Examples of Items per Group of Items and Operation*

Group / Item Type	Multiplication	Division
<b>Group of Small Numbers</b>		
CC	$3 \times \_ = 12$	$56 \div \_ = 8$
IC	$4 \times \_ = 31$	$26 \div \_ = 9$
II 1	$30 \times \_ = 6$	$7 \div \_ = 42$
II 2	$55 \times \_ = 8$	$3 \div \_ = 11$
Distractors	$0 \times \_ = 42$	$12 \div \_ = 0$
<b>Group of Decimal Numbers</b>		
CC	$3.1 \times \_ = 15.5$	$12.8 \div \_ = 3.2$
IC	$6.1 \times \_ = 17.2$	$7.5 \div \_ = 4.3$
II1	$18.3 \times \_ = 6.1$	$4.3 \div \_ = 8.6$
II2	$14.4 \times \_ = 3.1$	$3.2 \div \_ = 11.7$
Distractors	$0 \times \_ = 10.8$	$8.6 \div \_ = 0$
<b>Group of Large Numbers</b>		
CC	$6 \times \_ = 498$	$292 \div \_ = 4$
IC	$7 \times \_ = 384$	$735 \div \_ = 8$
II1	$438 \times \_ = 3$	$9 \div \_ = 657$
II2	$291 \times \_ = 4$	$6 \div \_ = 497$
Distractors	$0 \times \_ = 438$	$657 \div \_ = 0$

Combinations of the item characteristics produced four types of items, per group, that differed in their number-congruency and operation-congruency. First, there were items that were number-congruent and operation-congruent (e.g.,  $3 \times \_ = 12$ ); these items were called CC (abbreviation for Congruent/Congruent). In the Group of Small Numbers, the CC items were all part of the multiplication table, and would be familiar to the participants. The remaining item types were number-incongruent because substituting missing number symbols with a natural number would not make the equation true; thus, substituting only natural numbers would lead participants to the incorrect conclusion that the expression was not true. Second, some number-incongruent items were operation-congruent, in which multiplication produced bigger outcomes and division smaller outcomes; for example:  $4 \times \_ = 31$ . These items are called IC (abbreviation for Incongruent/Congruent).

The third and fourth remaining categories of items were both number-incongruent and operation-incongruent. For these items, the missing number was a non-natural number and the size of the results of the operations were not in-line with natural number arithmetic. Thus, both item types are II (abbreviation for Incongruent/Incongruent). In the first category of II items (named: II1) the missing number was a unit fraction. The II1 item number combinations may have appeared familiar to participants because they evoke number combinations that appear in the multiplication table, but these items had an inversion in the operand and the output. Thus, these items were false familiar because, for example,  $7 \div \_ = 42$  evokes '6' as missing number, but the correct answer is  $1/6$ ; and this appears familiar because  $42 \div 6 = 7$ . In the second category of II items (named: II2), there was a missing rational number smaller than one, which resulted in incongruent outcome size for each operation, i.e., smaller than the given operand numbers in cases of multiplication and larger in cases of division (e.g.,  $55 \times \_ = 8$ , in which case the unknown number is  $8/55$ , or  $0.1455$ ).

There were 36 items for each Group: 18 items per operation, with three items of each Type, plus 6 distractor items. Note that there were no items that were natural number-congruent and operation-incongruent. In this case, the operation would need to be the opposite (multiplication makes smaller; division makes bigger), but with a natural number missing, such items are mathematically impossible to create.

### Distractor Items

Since all experimental items had a number that would make the equation true, there were also distractor items that had no solution, and thus, for these items the correct response was that the equation could not be true. Distractors were arithmetic expressions with zero as an operator or outcome (see Table 1), since multiplication by zero always results in zero and division can never have zero as a quotient. Each group had 12 distractor items.

### Procedure

Participants completed the experiment in one 30-minute session in a classroom on a computer. Participants provided demographic information (i.e., age, gender) and then completed the experimental task. Stimuli were presented using E-prime. The items were presented in three blocks, with a break between each block. Each block included the four item types for one group. The order of the blocks was randomized.

Each equation was displayed with the question: *Can this be true? Y/N*. Participants responded by pressing *Y* for *Yes* and *N* for *No*. Participants were told that they did not need to find the missing number, only to respond whether a number exists that would make the equation true, as quickly and accurately as possible. Items were displayed until response and were preceded by a 500 ms fixation cross. Response time and accuracy were recorded for each trial.

### Predictions

We had several predictions for participants' response patterns to the four item types. If the participants' responses were based on mentally substituting natural numbers, there would be greater accuracy (Prediction 1a) and shorter reaction time (Prediction 1b) on the number-congruent/operation-congruent items (i.e., CC) than on the number-incongruent/operation-congruent items (i.e., IC). Additionally, if participants based their answers on the effect of operation, then there would be higher accuracy (Prediction 2a) and shorter reaction time (Prediction 2b) on operation-congruent items (i.e., the CC and IC items) than on operation-incongruent items (i.e., the II1 and II2 items). If participants were affected by the familiarity of the number combinations from the multiplication table, there would be higher accuracy (Prediction 3a) and shorter reaction time (Prediction 3b) on false familiar items (II1 items) than on the other number-incongruent/operation-incongruent items (II2 items), even though they are both number and operation-incongruent types of items.

Predictions 1 to 3 as described so far apply to all the three Groups of items. They are the core predictions of our study, as they specifically refer to the way the items are solved: by mentally substituting (specific) natural numbers or by judging the effect of the given operation. However, we further wanted to explore differential effects for the three Groups of items, which may further corroborate the underlying mechanisms. For instance, for the Group of Decimal Numbers and the Group of Large Numbers, we predict the effect of number-congruency will be less strong than in the Group of Small Numbers, because small number combinations are more easily memorized (Prediction 4). Especially for missing natural numbers (CC items), it is easier to determine

the missing number in the Group of Small numbers, and in some cases in the Group of Decimals (e.g., the missing number 5 for  $3.1 \times \_ = 15.5$ ) than in the Group of Large Numbers (e.g.  $6 \times \_ = 498$ ). Therefore, participants may tend to think less of a specific number for equations with decimal numbers and even less for equations with large numbers, than for equations with small numbers. Instead, for decimal and large number combinations, participants may count more on the operation-congruency effect (i.e., whether there are larger outcomes for multiplication and smaller for division). Thus, a stronger effect of operation-congruency in the Group of Large Numbers and the Group of Decimal Numbers would be expected, compared to the Group of Small Numbers (Prediction 5). For the same reasons, it would be expected that the effect of false familiarity would be stronger in the Group of Small Numbers than in the other groups (Prediction 6).

## Data Analysis

Distractor items were excluded from all analyses. Responses for multiplication and division items were analyzed separately, since the notions that multiplication makes bigger and division makes smaller may have differential effects. A separate analysis therefore provides a clearer picture of the phenomenon at hand for each operation. For each of the three groups, and again separately for multiplication and division items, we calculated outliers as trials that were more than three standard deviations above or below the average reaction time. Excluding distractor items, 18 of 3327 correct trials (0.5%) were excluded in total.

The Generalized Estimating of Equations (GEE) module in SPSS was used, with logistic regression to model accuracy (i.e., correct = 1 and incorrect = 0) and linear regression to model reaction time. The GEE approach accounts for the dependence of repeated measurements within subjects. In order to test the effect of number and operation-congruency the GEE analyses were applied to the accuracy rates, and also to reaction times, in the four types of items, for each one of the three Groups of items and each operation type. For both outcomes, Bonferroni-adjusted pairwise comparisons were applied in order to test possible statistically significant differences between item types. For accuracy rates, odds ratios were calculated for these pairwise comparisons, and for reaction times Cohen's *d* was used. To test for differences within each item type across the three groups, GEE analyses were also applied to the group analyses for each type. In this case, odd ratios were compared to account for differences in the effect of number and operation-congruency between the different groups of number combinations.

## Results

### Number and Operation-Congruency Effects on Accuracy Rates

#### The Effect of Multiplication for Each Group of Items

Table 2 shows participants' performance on all multiplication items, by item and group. GEE analysis of mean performance showed that in the Group of Small Numbers there was a significant effect of Type,  $\chi^2(3, N = 924) = 82.059, p < .001$ . Post-hoc Bonferroni-adjusted pairwise comparisons showed a statistically significant difference between CC and IC,  $p < .001, OR = 28.17, 95\% CI [9.62, 82.53]$ , CC and II1,  $p < .001, OR = 8.00, 95\% CI [2.67, 23.98]$ , and CC and II2,  $p < .001, OR = 29.33, 95\% CI [10.01, 85.95]$ , in which CC items had the highest accuracy. Estimated odds ratios suggest that the odds of answering a CC item correctly were about 28 times the odds of answering an IC item correctly and about 29 times the odds of answering an II2

item correctly. Both of these odds ratios are higher than the estimated odds ratio of answering a CC item correctly, versus II1 item. Taken together, these results indicate a large effect for number-congruency (in-line with Prediction 1a), which is stronger between tasks that appear unfamiliar. Participants performed higher in IC than in II2 items which, in line with Prediction 2a, indicates an effect for operation-congruency but this difference was not statistically significant,  $p = 1.000$ ,  $OR = 1.04$ , 95% CI [0.6, 1.82]. Not in line with Prediction 2a, participants performed statistically significantly higher on II1 items than on IC items,  $p < .001$ ,  $OR = 3.52$ , 95% CI [1.93, 6.41]. Finally, accuracy on II1 items was statistically significantly higher than on II2 items,  $p < .001$ ,  $OR = 3.67$ , 95% CI [2.01, 6.68], which indicate that the familiarity of the number combinations in the II1 items affected participants' responses (Prediction 3a). For small numbers, the effect of number-congruency is the largest, followed by the effect of familiarity. However, these results do not support a strong effect of operation-congruency for small numbers.

Table 2

*Accuracy Estimates in Multiplication for Each of the Four Types of Items, for Each Group of Items*

Group / Item Type	<i>M</i>	<i>SE</i>	95% Wald CI	
			<i>LL</i>	<i>UL</i>
<b>Group of Small Numbers</b>				
CC (e.g., $3 \times \_ = 12$ )	.96	.012	.93	.98
IC (e.g., $4 \times \_ = 31$ )	.46	.064	.34	.58
II1 (e.g., $30 \times \_ = 6$ )	.75	.044	.65	.82
II2 (e.g., $55 \times \_ = 8$ )	.45	.059	.34	.57
<b>Group of Decimal Numbers</b>				
CC (e.g., $3.1 \times \_ = 15.5$ )	.95	.015	.91	.97
IC (e.g., $6.1 \times \_ = 17.2$ )	.56	.046	.47	.65
II1 (e.g., $18.3 \times \_ = 6.1$ )	.74	.041	.65	.81
II2 (e.g., $14.4 \times \_ = 3.1$ )	.46	.049	.37	.56
<b>Group of Large Numbers</b>				
CC (e.g., $6 \times \_ = 498$ )	.83	.026	.78	.88
IC (e.g., $7 \times \_ = 384$ )	.69	.040	.60	.76
II1 (e.g., $438 \times \_ = 3$ )	.60	.047	.51	.69
II2 (e.g., $291 \times \_ = 4$ )	.45	.047	.36	.54

For the Group of Decimal Numbers, there was a statistically significant effect of Type  $\chi^2(3, N = 924) = 90.477$ ,  $p < .001$ . Pairwise comparisons showed a statistically significant difference between Types CC and IC,  $p < .001$ ,  $OR = 14.93$ , 95% CI [5.59, 39.86], CC and II1,  $p < .001$ ,  $OR = 6.68$ , 95% CI [2.45, 18.22], and CC and II2,  $p < .001$ ,  $OR = 22.3$ , 95% CI [8.36, 59.52], in which CC items always had the highest accuracy. Again, the estimated odds ratios suggest that the odds of answering a CC item correctly were about 14 times the odds of answering an IC item correctly and about 22 times the odds of answering an II2 item correctly. Both of these odds ratios are higher than the estimated odds ratio of answering a CC item correctly, versus a II1 item. Taken together, these results show an effect for number-congruency for decimal numbers (in line with Prediction 1a) which is stronger between tasks that do not appear familiar. Additionally, there was a statistically significant difference between IC and II2,  $p = .040$ ,  $OR = 1.49$ , 95% CI [0.86, 2.61], which indicates an effect for operation-congruency (Prediction 2a). There was also a statistically significant difference between IC and II1 in which, counter to Prediction 2a, performance on II1 items was higher than on IC items,  $p = .010$ ,  $OR = 2.24$ ,

95% CI [1.23, 4.06]. Accuracy in II1 items was statistically significantly higher than in II2 items,  $p < .001$ ,  $OR = 3.34$ , 95% CI [1.84, 6.06], which suggests quite an effect of familiarity for II1 items (Prediction 3a). In sum, for decimal numbers, number-congruency appears to have the largest effect, followed by familiarity, whereas operation-congruency appears to have the smallest effect.

For the Group of Large Numbers, there was a statistically significant effect of Type,  $\chi^2(3, N = 924) = 85.235$ ,  $p < .001$ . There was a statistically significant difference between CC and IC,  $p < .001$ ,  $OR = 2.19$ , 95% CI [1.12, 4.3], CC and II1,  $p < .001$ ,  $OR = 3.25$ , 95% CI [1.69, 6.28], and CC and II2,  $p < .001$ ,  $OR = 5.97$ , 95% CI [3.1, 11.47], in which CC items always had the highest accuracy. As above, these results suggest an effect of number-congruency (Prediction 1a) however, analysis of odd ratios showed less strong effects than in previous groups. The estimated odds ratios suggest that the odds of answering a CC item correctly were about 6 times the odds of answering an II2 item correctly and about 3 times the odds of answering an II1 item correctly, and both of these odds ratios are higher than the estimated odds ratio of answering a CC item correctly, versus a IC item. Finally, in-line with Prediction 2a, participants' performance on IC items was higher than on II1 and II2; the difference between IC and II2 was statistically significant,  $p < .001$ ,  $OR = 2.72$ , 95% CI [1.53, 4.85], but the difference between IC and II1 was not statistically significant,  $p = .762$ ,  $OR = 1.48$ , 95% CI [0.83, 2.66]. Additionally, accuracy on II1 items was higher than in II2 items,  $p < .001$ ,  $OR = 1.83$ , 95% CI [1.05, 3.21], which suggests an effect of familiarity (Prediction 3a). Accuracy rates for multiplication items by type and group are visually presented in the [Appendix](#), in Panel (a) of [Figure A1](#). In sum, for large numbers, the effect of number-congruency was largest, followed by the effect of operation-congruency, and the effect of familiarity was the smallest. However, the effect of number-congruency appeared smaller for large numbers than for small numbers and decimal numbers.

### The Effect of Division for Each Group of Items

[Table 3](#) shows participants' performance on division items, by item and group. For Small Numbers, there was a statistically significant effect of Type,  $\chi^2(3, N = 924) = 201.732$ ,  $p < .001$ . Bonferroni-adjusted pairwise comparisons showed a statistically significant difference between CC and IC,  $p < .001$ ,  $OR = 25.19$ , 95% CI [9.43, 67.28], CC and II1,  $p < .001$ ,  $OR = 15.55$ , 95% CI [5.82, 41.49], CC and II2,  $p < .001$ ,  $OR = 51.37$ , 95% CI [18.86, 139.89], in which CC items always had the highest accuracy. Estimated odds ratios suggest that the odds of answering a CC item correctly were about 25 times the odds of answering an IC item correctly and about 51 times the odds of answering an II2 item correctly. Both of these odds ratios are higher than the estimated odds ratio of answering a CC item correctly, versus a II1 item. These indicate an effect for number-congruency (in line with Prediction 1a) which is again stronger between tasks that appear unfamiliar. There was a statistically significant difference between IC and II2 items,  $p < .001$ ,  $OR = 2.04$ , 95% CI [1.13, 3.69], with higher performance on IC items which is partially in-line with Prediction 2a and indicates an effect of operation-congruency. On the other hand, there was no statistically significant differences between IC and II1 items,  $p = .284$ ,  $OR = 0.62$ , 95% CI [0.35, 1.08]. Finally, accuracy in II1 items was higher than in II2 items, and this difference was statistically significant,  $p < .001$ ,  $OR = 3.3$ , 95% CI [1.83, 5.97], (Prediction 3a). These results suggest that for small numbers, the effect of number-congruency was larger than the effect of familiarity, which was in turn larger than the effect of operation-congruency.

Table 3

Accuracy Estimates in Division for Each of the Four Types of Items, for Each Group of Items

Group / Item Type	<i>M</i>	<i>SE</i>	95% Wald CI	
			<i>LL</i>	<i>UL</i>
<b>Group of Small Numbers</b>				
CC (e.g., 56 ÷ _ = 8)	.95	.014	.91	.97
IC (e.g., 26 ÷ _ = 9)	.43	.047	.35	.53
II1 (e.g., 7 ÷ _ = 42)	.55	.050	.45	.64
II2 (e.g., 3 ÷ _ = 11)	.27	.047	.19	.37
<b>Group of Decimal Numbers</b>				
CC (e.g., 12.8 ÷ _ = 3.2)	.93	.017	.89	.96
IC (e.g., 7.5 ÷ _ = 4.3)	.49	.050	.39	.58
II1 (e.g., 4.3 ÷ _ = 8.6)	.64	.046	.54	.72
II2 (e.g., 3.2 ÷ _ = 11.7)	.39	.046	.30	.48
<b>Group of Large Numbers</b>				
CC (e.g., 292 ÷ _ = 4)	.93	.028	.72	.83
IC (e.g., 735 ÷ _ = 8)	.49	.046	.44	.62
II1 (e.g., 9 ÷ _ = 657)	.64	.045	.36	.53
II2 (e.g., 6 ÷ _ = 497)	.32	.045	.24	.42

For Decimal Numbers, there was a statistically significant effect of Type,  $\chi^2(3, N = 924) = 132.866, p < .001$ . In-line with Predictions 1a and 2a, there was a statistically significant difference between CC and IC,  $p < .001$ ,  $OR = 13.83$ , 95% CI [5.84, 32.76], CC and II1,  $p < .001$ ,  $OR = 7.47$ , 95% CI [3.13, 17.84], CC and II2,  $p < .001$ ,  $OR = 20.78$ , 95% CI [8.73, 49.45], in which CC items had the highest accuracy. Estimated odds ratios suggest that the odds of answering a CC item correctly were about 13 times the odds of answering an IC item correctly and about 20 times the odds of answering an II2 item correctly. Both of these odds ratios are higher than the estimated odds ratio of answering a CC item correctly, versus a II1 item. These results show an effect of number-congruency (in line with Prediction 1a) which is stronger between unfamiliar tasks. There was a statistically significant difference between IC and II2 items,  $p = .008$ ,  $OR = 1.50$ , 95% CI [0.86, 2.63], reflects an effect of operation-congruency, but the difference between IC and II1 did not reach statistical significance,  $p = .075$ ,  $OR = 0.54$ , 95% CI [0.31, 0.95], providing only partial support for Prediction 2a. Finally, accuracy in II1 items was higher than in II2 items, and this difference was statistically significant,  $p < .001$ ,  $OR = 2.78$ , 95% CI [1.57, 4.93], (Prediction 3a). For decimal numbers, the effect of number-congruency was the largest, followed by the effect of familiarity, with the smallest effects and limited support for operation-congruency.

For the Group of Large Numbers, there was a statistically significant effect of Type,  $\chi^2(3, N = 924) = 101.611, p < .001$ , for division items. There was a statistically significant difference between CC and Types IC,  $p < .001$ ,  $OR = 13.83$ , 95% CI [5.84, 32.76], CC and II1,  $p < .001$ ,  $OR = 9.55$ , 95% CI [4.05, 22.49], CC and II2,  $p < .001$ ,  $OR = 28.23$ , 95% CI [11.76, 67.76], in which CC items had the highest accuracy. Estimated odds ratios suggest that the odds of answering a CC item correctly were about 14 times the odds of answering an IC item correctly and about 28 times the odds of answering an II2 item correctly. Both of these odds ratios are higher than the estimated odds ratio of answering a CC item correctly, versus a II1 item. These results show an effect of number-congruency (in-line with Prediction 1a), especially for items that appear unfamiliar. Additionally, there was a significant difference between IC and II2,  $p < .001$ ,  $OR = 2.04$ , 95% CI [1.15, 3.63], with higher performance

in the IC items, which refers to a difference concerning operation-congruency (in-line with Prediction 2a). There was no statistically significant difference between IC and II1,  $p = .724$   $OR = 1.38$ , 95% CI [0.79, 2.40]. Finally, accuracy in II1 items was higher than in II2 items, and this difference was statistically significant,  $p = .003$ ,  $OR = 3.78$ , 95% CI [2.10, 6.79], (Prediction 3a). For large numbers, number-congruency had the largest effect, with smaller effects for familiarity and operation-congruency. Accuracy rates for division items by type and group are visually presented in the [Appendix](#), in Panel (b) of [Figure A1](#).

In sum, the above results supported Predictions 1a about the effect of number-congruency in participants' accuracy rates, since there were statistically significantly higher performances in the number-congruent items (i.e., CC), than in the number-incongruent items (i.e., IC, II1, II2) items, in all groups of items, in multiplication and also in division. Results suggest mixed support for Prediction 2a that participants would tend to base their answers on the effect of operation and there would be higher accuracy for operation-congruent than operation-incongruent items. For all groups of items there were higher accuracy rates for IC items than II2 items in both operations, which supports Prediction 2a about the effect of operation-congruency. However, according to Prediction 2a, IC items should also have higher accuracy than II1 items, but this was only the case in the Group of Large Numbers in multiplication. Lower accuracies in operation-congruent IC items than in the operation-incongruent II1 items suggests a familiarity effect on participants' evaluations about the validity of the multiplication and division equations. In-line with this interpretation, results support Prediction 3a, that there would be higher accuracy for false familiar items, with higher accuracy on II1 items than on II2 items, in which in both cases the items are operation-incongruent. The comparison of odd ratios showed that, for both operations, the number-congruency effect was consistently the largest, in relation to the operation-congruency effect or the effect of familiarity, and these effects were larger for the groups of small and decimal numbers than for the group of large numbers.

To further explore the above results, a GEE model was applied to accuracy rates across the three Groups for selected pairs of item types (i.e., comparing CC and IC; IC and II1; IC and II2; and II1 and II2). Comparing the odd ratios provides indications about the possible differences between the number and operation-congruency effects, in each Group of items.

## The NNB Effect Between the Three Groups of Items

### The Number-Congruency Effect Between the Groups

To test the main effect of number-congruency and the interaction of number-congruency and type, a GEE analysis was applied to the CC and IC Items, for the three Groups. For multiplication, there was a statistically significant interaction effect of group and number-congruency,  $\chi^2(5, N = 1386) = 98.767$ ,  $p < .001$ . Performance was higher in CC than IC items in the Group of Small Numbers,  $p < .001$ ,  $OR = 55.26$ , 95% CI [12.91, 236.51], in the Group of Decimal Numbers,  $p < .001$ ,  $OR = 18.11$ , 95% CI [6.18, 53.08], and in the Group of Large Numbers,  $p < .001$ ,  $OR = 2.40$ , 95% CI [1.23, 4.69]. The odd ratios showed that, in-line with Prediction 4, the effect of number-congruency was larger in the Group of Small Numbers than in the other groups.

Results for division were similar, in which there was also a statistically significant interaction effect of group and number-congruency,  $\chi^2(5, N = 1386) = 113.682$ ,  $p < .001$ . Again, performance in CC items was higher than in IC items in the Group of Small Numbers,  $p < .001$ ,  $OR = 26.00$ , 95% CI [8.88, 76.13], in the Group of Decimal Numbers,  $p < .001$ ,  $OR = 9.80$ , 95% CI [4.30, 22.30], and in the Group of Large Numbers,  $p < .001$ ,  $OR =$

3.73, 95% CI [1.96, 7.10]. The odds ratios showed again that the effect of number-congruency was larger in the Group of Small than in the other groups (in-line with Prediction 4).

### The Operation-Congruency Effect Between the Groups of Items

The same analysis was applied to the IC and II1 items to test the main effect of operation-congruency and the interaction of operation-congruency and type, for the three Groups. For multiplication, there was a statistically significant interaction effect of group and operation-congruency,  $\chi^2(5, N = 1386) = 43.436, p < .001$ . Accuracy in II1 items was higher than in IC items in the Group of Small Numbers,  $p < .001, OR = 3.34, 95\% CI [1.84, 6.06]$ , and in the Group of Decimal Numbers but this difference was not statistically significant,  $p = .061, OR = 2.11, 95\% CI [1.15, 3.88]$ . In the Group of Large Numbers, accuracy in IC items was higher than in II1 items but again this difference was not statistically significant,  $p = 1.000, OR = 1.19, 95\% CI [0.67, 2.10]$ . The odds ratio showed that, in-line with Prediction 5, the effect of operation-congruency was largest in the Group of Small Numbers in which performance on II1 items was higher than in IC items. For division, participants' performance on II1 items was always higher than for IC items, and the odds ratios showed the same pattern as in multiplication; however, the interaction effect of group and operation-congruency was not statistically significant,  $\chi^2(5, N = 1386) = 11.036, p = .051$ .

Additionally, the main effect of operation-congruency and the interaction of operation-congruency and type was tested in the IC and II2 items, for the three Groups. For multiplication, there was a statistically significant interaction effect of group and operation-congruency,  $\chi^2(5, N = 1386) = 36.075, p < .001$ , with higher performance on IC items than performance on II2 items in the Group of Large Numbers,  $p < .001, OR = 2.57, 95\% CI [1.45, 4.56]$ , where the differences were statistically significant, and also in the Group of Decimal Numbers,  $p = 1.000, OR = 1.38, 95\% CI [0.79, 2.42]$ , and in the Group of Small Numbers,  $p = 1.000, OR = 1.04, 95\% CI [0.60, 1.81]$ , where the differences were not statistically significant. The odds ratios showed that the effect of operation-congruency is larger in the Group of Large Numbers, compared to its effect in the other groups (in-line with Prediction 5).

For division, there was also a statistically significant interaction effect of group and operation-congruency,  $\chi^2(5, N = 1386) = 55.173, p < .001$ , with higher performance on IC items than on II2 items in the Group of Small Numbers,  $p < .001, OR = 3.14, 95\% CI [1.70, 5.81]$ , in the Group of Decimal Numbers,  $p < .001, OR = 2.43, 95\% CI [1.36, 4.34]$ , and in the Group of Large Numbers,  $p = .022, OR = 2.07, 95\% CI [1.11, 3.80]$ . However, counter to Prediction 5, the odds ratios showed that the effect of operation-congruency is larger in the Group of Small Numbers than in the other groups.

### The Effect of Item Familiarity Between the Groups of Items

The same analysis was also conducted on the II1 and II2 items, to test the effect of familiarity with the items that appear in the multiplication table, for the three Groups. For multiplication, there was a statistically significant interaction effect of group and familiarity,  $\chi^2(5, N = 2079) = 123.435, p < .001$ . Accuracy in II1 items was higher than in II2 items in Group of Small Numbers,  $p < .001, OR = 5.88, 95\% CI [3.18, 10.89]$ , in the Group of Decimal Numbers,  $p < .001, OR = 4.1, 95\% CI [2.25, 7.45]$ , and also in the Group of Large Numbers.  $p < .001, OR = 2.26, 95\% CI [1.28, 3.99]$ . The odds ratio showed that, in-line with Prediction 6, the effect of familiarity is larger in the Group of Small Numbers than in the Group of Decimal Numbers and is smallest in the Group of Large Numbers.

For division, the odds ratios showed almost the same pattern. There was also a statistically significant interaction effect of group and familiarity,  $\chi^2(5, N = 2079) = 32.196, p < .001$ . Accuracy in II1 items was higher than in II2 items in Group of Small Numbers,  $p = .004$ ,  $OR = 2.01$ , 95% CI [1.14, 3.55], in the Group of Decimal Numbers,  $p < .001$ ,  $OR = 2.25$ , 95% CI [1.28, 3.96], and in the Group of Large Numbers,  $p = .194$ ,  $OR = 1.51$ , 95% CI [0.86, 2.67], in which the differences were not statistically significant. The odds ratio showed that the effect of familiarity was similar in magnitude in the Group of Small Numbers and the Group of Decimal Numbers and smaller in the Group of Large Numbers.

In summary, for both multiplication and division, the effect of number-congruency was larger in the Group of Small Numbers than the Group of Decimal Numbers and the Group of Large Numbers, results that support Prediction 4 that the effect of number-congruency will be less strong than in the Group of Small Numbers, because small number combinations are more easily memorized. Higher accuracy rates on the IC than on the II2 items showed an effect of operation-congruency which was larger in the Group of Large Numbers than in the Group of Decimal Numbers and in the Group of Small Numbers, in multiplication, but not in division. These results support Prediction 5 that a stronger effect of operation-congruency would be expected in the Group of Large Numbers and the Group of Decimal Numbers compared to the Group of Small Numbers. Interestingly, higher accuracy on IC than II1 items in multiplication also showed an operation-congruency effect that was larger in the Group of Large Numbers than in the Group of Decimal Numbers and the Group of Small Numbers, where accuracy was higher in II1 than in the IC items. Again, these results may indicate that the effect of familiarity (i.e., the II1 items) is larger in small number combinations than in decimal and large number combinations. In support of this interpretation, in-line with Prediction 6, the effect of familiarity was larger in the Group of Small Numbers than the Group of Decimal Numbers and the Group of Large Numbers as shown in the higher accuracy rates in II1 than in II2 items for both operations.

## Analysis of the Reaction Times

### Results for Reaction Time for Multiplication Items

Table 4 shows participants' mean reaction times for multiplication items, for each group and type. In the Group of Small Numbers, there was a statistically significant effect of Type,  $\chi^2(3, N = 603) = 24.007, p < .001$ . In-line with Prediction 1b the participants responded to CC items faster than to IC items,  $p = .023, d = 0.39$ , II1 items,  $p = .001, d = 0.45$ , and II2 items,  $p < .001, d = 0.47$ . There was no statistically significant difference between IC and II1 items,  $p = 1.000, d = 0.057$ , IC and II2 items,  $p = 1.000, d = 0.072$ , or II1 and II2 items,  $p = 1.000, d = 0.015$  (Prediction 3b). These suggest a medium effect of number-congruency for small numbers.

In the Group of Decimal Numbers there was a statistically significant effect of Type,  $\chi^2(3, N = 616) = 12.756, p = .005$ . In-line with Prediction 1b, there were statistically significant differences between mean response time on the CC and II1 items,  $p = .007, d = 0.22$ , and between the CC and II2 items,  $p = .010, d = 0.38$ . There was not a statistically significant difference in reaction time between the CC and IC items,  $p = 1.000, d = 0.057$ . Additionally, mean reaction time was statistically significantly faster on IC items than on II2 items,  $p = .029, d = 0.41$ , but there were no statistically significant differences between mean reaction time on the IC and II1 items,  $p = .253, d = 0.25$ , (Prediction 2b), or between the II1 and II2 items,  $p = 1.000, d = 0.16$  (Prediction 3b). Together, these results suggest small to medium effects for number-congruency and support for medium effects of operation-congruency for decimal numbers.

Table 4

Mean Reaction Times of Correctly Solved Multiplication Items by Group and Item Type

Group / Item Type	<i>M</i>	<i>SE</i>	95% Wald CI	
			<i>LL</i>	<i>UL</i>
<b>Group of Small Numbers</b>				
CC (e.g., 3 × _ = 12)	1901.98	139.811	1627.96	2176.00
IC (e.g., 4 × _ = 31)	2699.41	237.433	2234.05	3164.77
II1 (e.g., 30 × _ = 6)	2814.38	208.210	2406.30	3222.46
II2 (e.g., 55 × _ = 8)	2844.59	237.221	2379.64	3309.53
<b>Group of Decimal Numbers</b>				
CC (e.g., 3.1 × _ = 15.5)	2823.10	204.423	2422.44	3223.76
IC (e.g., 6.1 × _ = 17.2)	2752.19	261.682	2239.30	3265.08
II1 (e.g., 18.3 × _ = 6.1)	3385.88	268.082	2860.45	3911.32
II2 (e.g., 14.4 × _ = 3.1)	3783.57	323.882	3148.77	4418.36
<b>Group of Large Numbers</b>				
CC (e.g., 6 × _ = 498)	3238.68	257.427	2734.13	3743.23
IC (e.g., 7 × _ = 384)	2911.23	236.297	2448.10	3374.37
II1 (e.g., 438 × _ = 3)	3950.17	303.911	3354.51	4545.82
II2 (e.g., 291 × _ = 4)	3663.03	296.369	3082.16	4243.91

In the Group of Large Numbers, there was a statistically significant effect of Type,  $\chi^2(3, N = 553) = 18.481, p < .001$ . However, results only partially supported Prediction 1b. There were statistically significant differences in mean reaction times between CC and II1 items, with faster responses on CC items,  $p = .007, d = 0.25$  (Prediction 1b). There was no statistically significant difference in mean reaction time between the CC and IC items,  $p = 1.000, d = 0.11$ . Additionally, there were statistically significant differences in mean reaction times between IC and II1 items,  $p = .001, d = 0.37$ , and also between IC and II2 items,  $p = .030, d = 0.26$ , with faster responses on IC items (Prediction 2b). As with the other two groups of numbers, differences in mean reaction time between II1 and II2 items were not statistically significant,  $p = 1.000, d = 0.10$  (Prediction 3b). Together, these results suggest limited support for a small effect of number-congruency for large numbers and small to medium effects of operation-congruency. Notably, there is no evidence of a familiarity effect; we expect no effect, since the numbers are outside of the multiplication table. Mean reaction time for multiplication items by type and group are presented in the Appendix, in Panel (a) of Figure A2.

### Results for Reaction Time for Division Items

Table 5 shows participants' mean performance by group and item. There was no statistically significant effect of Type in the Group of Small Numbers,  $\chi^2(3, N = 490) = 2.921, p = .404$ , or in the Group of Decimal Numbers,  $\chi^2(3, N = 546) = 0.914, p = .822$ . However, there was a statistically significant effect of Type in the Group of Large Numbers,  $\chi^2(3, N = 458) = 9.565, p = .023$ . There was a statistically significant difference only between mean reaction time in IC and II1 items ( $p = .013, d = 0.32$ ), with faster reaction times on the IC items than on the II1 items. This result partly supports Prediction 2b. Mean reaction time for division items by type and group are presented in the Appendix, in Panel (b) of Figure A2.

Table 5

Mean Reaction Times of Correctly Solved Division Items for the Three Number Groups

Group / Item type	<i>M</i>	<i>SE</i>	95% Wald CI	
			<i>LL</i>	<i>UL</i>
<b>Group of Small Numbers</b>				
CC (e.g., $56 \div \_ = 8$ )	2390.05	128.475	2138.24	2641.85
IC (e.g., $26 \div \_ = 9$ )	2386.81	233.145	1929.85	2843.77
II1 (e.g., $7 \div \_ = 42$ )	2702.49	199.940	2310.62	3094.37
II2 (e.g., $3 \div \_ = 11$ )	2877.80	393.923	2105.72	3649.87
<b>Group of Decimal Numbers</b>				
CC (e.g., $12.8 \div \_ = 3.2$ )	2963.79	189.703	2591.98	3335.60
IC (e.g., $7.5 \div \_ = 4.3$ )	3054.95	228.722	2606.66	3503.24
II1 (e.g., $4.3 \div \_ = 8.6$ )	3050.05	230.782	2597.72	3502.37
II2 (e.g., $3.2 \div \_ = 11.7$ )	3161.36	267.176	2637.70	3685.01
<b>Group of Large Numbers</b>				
CC (e.g., $292 \div \_ = 4$ )	3587.94	272.445	3053.95	4121.92
IC (e.g., $735 \div \_ = 8$ )	2889.53	249.229	2401.05	3378.01
II1 (e.g., $9 \div \_ = 657$ )	3841.23	285.222	3282.20	4400.25
II2 (e.g., $6 \div \_ = 497$ )	3560.87	323.871	2926.10	4195.65

In sum, in most cases, in-line with Prediction 1b, that responses on the number-congruent/operation-congruent items would be quicker, the participants responded statistically significantly faster on the CC items than on II1 and II2 items. Also, in-line with Prediction 2b, that there would be shorter reaction time on operation-congruent items participants responded faster on IC items than on II1 and II2 items. In some cases, the participants responded faster on IC than on CC items, however these differences were not statistically significant. Also, the participants' mean reaction time was in most cases faster in on false familiar items II1 than on II2 items (in-line with Prediction 3), however these differences were again not statistically significant. The patterns presented above were clear only in multiplication items in the Group of Small Numbers, and were less clear in the Group of decimal Numbers and the Group of Large Numbers, and almost absent in all groups of division items.

## Discussion

This study examined the effect of the NNB on operations between given and missing numbers, using accuracy and response time measurement to extend previous findings that suggest a dual effect of the NNB in arithmetic operations between missing numbers (Christou, 2015a, 2015b). The main hypothesis of the study was that the NNB affects evaluations about the validity of equations that present arithmetic operations between given and missing numbers in two main ways: a) a tendency to think that missing numbers are natural numbers; and b) a tendency to associate each operation with specific results independently of the numbers involved in the operations, i.e., larger results than the operand in multiplication and smaller results in division.

Based on the above hypotheses, the participants were expected to evaluate the validity of given equations (e.g.,  $7 \div \_ = 42$ ) using two main strategies: a) to mentally substitute different numbers to test the validity of the equations and b) to compare the operand and result. To test these assumptions, four types of equations were designed to be congruent or incongruent with the intuitive beliefs about the missing numbers and the size

of the results of each operation, and were tested using three groups of numbers (i.e., small natural numbers, decimal numbers, large natural numbers). Within the different number groups, the familiarity effect was tested, using number combinations that appear in the multiplication table or seem to appear in the multiplication table, i.e., false familiarity).

Results supported the first aspect of the NNB, showing a large effect of number-congruency on participants' evaluations for all groups of items, with higher accuracy rates on those items with missing natural numbers (i.e., the number-congruent, CC items), compared to those items that falsified this belief (i.e., all other items). These results support the main prediction of the study. The effect of number-congruency was larger in the Group of Small Number items than in the Group of Decimal Numbers and the Group of Large numbers. This suggests that participants may mentally test the effect of operations by substituting specific numbers for missing numbers, which under the influence of the NNB, are mostly natural numbers; a finding that further supports previous research (Christou, 2015a, 2015b; Van Hoof et al., 2015). This strategy is also easier to apply effectively in the case of small number combinations than in other categories of tasks. Importantly, it may appear that these items preclude our ability to disentangle the mechanisms of arithmetic fact retrieval from the NNB, if students try to find the missing number (presumably by substituting natural numbers). From our perspective, however, these two mechanisms may not be separable. Multiplication fact retrieval relies on memorization of the multiplication table, which comprises natural number arithmetic. The emphasis on developing fact fluency for natural number arithmetic throughout primary school years are among the factors that may strengthen the NNB phenomenon.

Results supported the second aspect of the NNB, an effect of operation-congruency on participants' evaluations, more clearly for multiplication than division. Specifically, the results showed higher accuracy rates for items that were in-line with participants' intuitive beliefs about the size of the results of multiplication and division (i.e., the operation-congruent, IC items), comparing with their evaluations in those items that falsified these beliefs when those items were not familiar (i.e., the operation-incongruent, II2 items). This is in-line with the prediction regarding the second part of the dual effect hypothesis, and with previous research that reported the *multiplication makes bigger* misconception (Fischbein et al., 1985; Greer, 1994; Onslow, 1990; Vamvakoussi et al., 2012, 2013; Van Hoof et al., 2015). For accuracy rates, results for multiplication items supported our prediction that the operation-congruency effect was larger in the Group of Large Number items and the Group of Decimal Numbers than in the Group of Small numbers. This shows that when strategies such as finding the missing number or recognizing number facts from the multiplication table are more difficult to apply, participants base their evaluations on their intuitions about the size of the results expected from each arithmetic operation. However, for division items, the operation-congruency effect was larger in the Group of Small Number items than in the Group of Decimal Numbers and the Group of Large numbers. This result requires further investigation, such as by coupling accuracy and reaction time data with participant interviews about their problem-solving strategies.

Further support for the notion that participants may base their evaluations on trying to find the missing number when this is possible was suggested with higher accuracy rates on operation-incongruent items that appeared familiar because the missing number was a unit fraction (i.e., II1 items), than operation-congruent items that were in-line with intuitive beliefs about the size of the results of operation (i.e., IC items), in all cases except in large number multiplication items; a result counter to our predictions. Most often these differences were not statistically significant, however they suggest that in cases where the tasks are operation-incongruent and

appear familiar, participants tended to respond even higher than in those tasks where the size of the results of the operations were in-line with participants' intuitive beliefs. In those cases, the effect of operation-congruency was stronger in the Group of Small Numbers than in the Group of Decimal and Large Numbers, both in multiplication and division. These results further support the above interpretation, that participants may base their evaluations on number substitution or task familiarity, rather than on the effect of the operation, and this is a more effective strategy in the case of small numbers.

Statistically significant differences between accuracy rates on items that falsified intuitions about number and operation-congruency (i.e., the II1 and the II2 items) further support this interpretation. As was predicted, accuracy was higher on items that appeared familiar (i.e., the II1 items) than on items that did not (i.e., the II2 items), for both multiplication and division. Again, this effect of false familiarity was stronger in the Group of Small Numbers than the other two groups, for both operations, however the effect sizes were smaller than the number congruency effect.

For reaction time, only multiplication items with small numbers supported the above interpretations; results were less clear for other multiplication items and groups, and for all division items and groups. In most cases, the participants responded statistically significantly faster on number-congruent and operation-congruent (i.e., CC) items than on the number-incongruent and operation-incongruent (i.e., the II1 and the II2) items. Also, as it was predicted, participants responded faster on items that were aligned with their intuitions about the size of the result of the operations (i.e., the IC items) and slower on items that falsified these intuitions (i.e., the II1 and the II2 items). There was also a familiarity effect for items that falsified intuitions about the effect of operations. In line with our predictions, participants' responses were most often faster when items appeared familiar (i.e., II1 items) than when they did not (i.e., II2 items). Interestingly, for items that aligned with intuitions about the size of the results of operations, in most cases participants responded slower when a natural number was missing (i.e., on the CC items) than when a non-natural number was missing (i.e., the IC items). These differences were not statistically significant; however, it is possible that there was not sufficient statistical power to detect these effects, due to large variation in reaction times overall. Speculatively, differences may suggest that participants spend time to confirm that a specific natural number is missing, yet do not when the missing number is not natural. This hypothesis could be tested in subsequent studies.

Overall, the results of the study support the dual effect of the NNB in arithmetic operations between missing numbers, providing further empirical support to previous findings (Christou, 2015a, 2015b; Obersteiner et al., 2016; Vamvakoussi et al., 2012, 2013; Van Hoof et al., 2015). Results also suggest that when small natural number combinations are presented with missing numbers, participants are inclined to mentally substitute natural numbers to decide whether the expression can be true, and thus they are more susceptible to number-congruency than operation-congruency effects, and to familiarity effects, especially in multiplication. However, when it is more difficult to trace the missing natural number, and the effect of familiarity is less strong, such as with decimal number or large number combinations, students base their evaluations on intuitions about the size of results from multiplication and division, as a more effective strategy. This tendency appeared when natural numbers were missing and when unit fractions were missing. The latter may have created an effect of false familiarity with multiplication facts, suggesting that students may directly retrieve facts from memory, as suggested in previous studies (Krueger, 1986; Krueger & Halford, 1984; Stazyk et al., 1982).

## Limitations

To focus on testing the dual effect of the NNB, this study maximized experimental items to examine number- and operation-congruency effects, and had relatively fewer distractor items (since distractor items must contain zero as an operand or result). As a result, to answer correctly, participants needed to indicate more often that the expression was possible than impossible. However, given the observed accuracy rates, participants evaluated many of the expressions to be impossible, suggesting item imbalance was not an issue.

Another potential limitation was that students could still respond using general insights about the solvability of linear equations. There may be contexts in which this strategy of applying a general insight (understanding that all expressions are possible as long as they don't involve a 0) is frequently relied on. For instance, [Obersteiner et al. \(2016\)](#) observed this in mathematical experts. In our study, missing number symbols rather than literal symbols were used to specifically discourage this strategy. Use of the general insight strategy in our current study seems unlikely; there were not ceiling effects, systematic fast reaction times on all items, or indications that participants figured out this strategy during the course of the experiment. Specifically, only one student gave the correct response for all items, and only 10% of participants scored higher than 95. From our perspective, these results show that even if some students used either of the two strategies (i.e., either solved the equations, or evoked knowledge on equation solvability) they did not do it systematically. This implies that although they presumably recognized that knowledge of equations was relevant to the tasks at hand, they failed to use this knowledge in all items. This could be interpreted as an instance where the intuitive, "biased" response ([Leron & Hazzan, 2006](#)) overrode the analytic, perhaps because students felt more confident about it. Future studies that incorporate interviews could shed more light on the actual strategies that the participants tend to use in each category of items, as the present study did not collect strategy information.

Finally, a limitation of this research approach generally is that it involves rating symbolic equations only in a purely symbolic form. Performance on mathematical tasks can be highly context dependent (e.g., [Saxe, 2015](#)), and so can knowledge of mathematical principles ([Prather & Alibali, 2008](#)). Along this line, someone could argue that thinking of the unknown number as a decimal number, for example, in a context where all the other numbers in the equation are natural, is more difficult than when decimal numbers would appear in the same equation. Indeed, in the present study, the accuracy rates on the tasks that were number incongruent but operation congruent were slightly higher in the group of decimal numbers than in the group of small natural numbers. However, since participants are exposed to natural and decimal numbers throughout the experiment, this potential difference in difficulty should be eliminated or greatly reduced for any specific equation. Additionally, from the NNB perspective that we endorse, such an effect of context would further support our main claim about an intuitive/analytic distinction between the cognitive processes that underlie reasoning with arithmetic operations. In other words, no context effect should appear if participants' responses were not affected by the NNB, considering that the participants have been exposed to decimal numbers and fractions for many years throughout schooling. Support for this position comes from previous studies which have shown that the NNB may affect how students apply different properties to the different symbolic representations of rational numbers, i.e., when they appear as decimals or as fractions (for a detailed discussion see [Vamvakoussi & Vosniadou 2010](#)). However, participants may still perform differently if tasks are presented in verbal or problem-solving form. As such, the potentially context-dependent nature of performance on these tasks requires further examination. Also, further studies should test whether the reported effects could be education dependent, by testing

how performance on the tasks might differ between adults with more and less formal education or with younger students.

## Implications

The results of the present study provide valuable information about the cognitive processes that underlie reasoning with arithmetic operations and the role of prior natural number knowledge in these processes. Results further support that the NNB contributes to the effect of number size on students' evaluations of the validity of multiplication and division equations, which further suggests that classroom instructors will need to address the NNB. However, addressing the NNB in instruction is a complex endeavor that requires remedial and anticipation approaches (Vamvakoussi et al., 2018; Van Dooren et al., 2015).

As an initial step, teachers should learn that students hold intuitive ideas about various mathematical topics, including the size and the type of the results of operations, since preservice teachers are insufficiently aware of such issues (Depaepe et al., 2015; Depaepe et al., 2018). Students should also learn about their intuitions, since these beliefs are often implicit and not under their conscious control (Fischbein, 1987). However, merely challenging students' erroneous beliefs, without interventions that raise awareness about the discrepancy between incorrect beliefs and the mathematically correct perspective, is not enough for students to abandon their alternative conceptions and accept the mathematically correct knowledge (Merenluoto & Lehtinen, 2004).

Further, merely raising awareness may not be enough. For example, students who can verbalize the fact that non-natural numbers can be substituted for missing number symbols or literal symbols may still substitute natural numbers only, even after hints from interviewers (Christou & Vosniadou, 2012; Van Hoof et al., 2015). Along the same line, merely changing the missing number symbol to  $x$  in the given equations (rather than the ' $\_$ ' that was used), would be unlikely to solve the NNB issue. That is because by following an equation-solving process (mentally or not), the participants may determine the non-natural missing number (i.e., the solution). This, however, would draw on cognitive processes different from students' intuitive responses that are affected by the NNB. Therefore, as has been argued before (Christou, 2015b; Dimitrakopoulou & Christou, 2018; Vamvakoussi et al., 2013), for students to overcome their intuitions about the effects of arithmetic operations, they need explicit strategies to inhibit natural number knowledge interference (Moutier & Houdé, 2003; Roell et al., 2017, 2019; Van Dooren & Inglis, 2015). One inhibition strategy is to always try with at least one non-natural number—a negative or a positive rational number smaller than 1—in cases of missing number tasks, or to always recall that *multiplication does not always make bigger*.

Lastly, the design of the items presented in the study may be fruitfully used and empirically tested as educational materials that could illuminate and falsify students' intuitive beliefs about the size of the results of operations. These items could be used in constructivist teaching environments to raise students' awareness about the familiarity of multiplication facts due to natural number calculations. Students can learn that the tendency to think that calculations between missing numbers hold only between natural numbers could create further constraints on successfully completing mathematical tasks, such as solving equations.

## Funding

The authors are members of the Centre for Innovative Research "Conceptual Change" supported by the European Association for Research on Learning and Instruction (EARLI).

## Competing Interests

The authors have declared that no competing interests exist.

## Acknowledgments

The authors have no support to report.

## Previously Presented

Results from this research were previously presented at several conferences:

Christou, K. P., Pollack, C., Vannuten, S., Van Hoof, J., & Van Dooren, W. (2016, June). New insights on natural number bias in arithmetic operations. In K. P. Christou & H. Eshach (Eds.), *Conceptual change meets other disciplines – Abstracts of the 10<sup>th</sup> International Conference on Conceptual Change* (p. 72). Florina, Greece: UOWM.

Pollack, C., Christou, K. P., Vannuten, S., Van Hoof, J., & Van Dooren, W. (2017). Natural number bias when reasoning with rational numbers and unknown operands. In *The Book of Abstracts and Extended Summaries of the 17th Biennial Conference of the European Association for Research on Learning and Instruction* (online edition). Tampere, Finland.

Van Dooren, W., Pollack, C., Vannuten, S., Van Hoof, J., & Christou, K. P. (2018) Natural number bias when reasoning about the effect of operations. In *The Book of Abstracts of the 11th International Conference on Conceptual Change – Epistemic Cognition & Conceptual Change* (p. 97), Klagenfurt, Austria: University of Klagenfurt.

## References

- Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, 44, 75-106. [https://doi.org/10.1016/0010-0277\(92\)90051-I](https://doi.org/10.1016/0010-0277(92)90051-I)
- Ashcraft, M. H., & Battaglia, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Learning and Memory*, 4(5), 527-538. <https://doi.org/10.1037/0278-7393.4.5.527>
- Bailey, D. H., Hoard, M. K., Nugent, L., & Geary, D. C. (2012). Competence with fractions predicts gains in mathematics achievement. *Journal of Experimental Child Psychology*, 113, 447-455. <https://doi.org/10.1016/j.jecp.2012.06.004>
- Behr, M., Harel, G., Post, T., & Lesh, R. (1994). Units of quantity: A conceptual basis common to additive and multiplicative structures. In G. Harrel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 121-176). Albany, NY, USA: State University of New York Press.
- Bell, A., Swan, M., & Taylor, G. (1981). Choice of operation in verbal problems with decimal numbers. *Educational Studies in Mathematics*, 12, 399-420. <https://doi.org/10.1007/BF00308139>

- Carpenter, T., Fennema, E., & Romberg, T. (1993). *Rational numbers: An integration of research*. Hillsdale, NJ, USA: Erlbaum.
- Christou, K. P. (2015a). The dual aspect of natural number bias in arithmetic operations. *Mediterranean Journal for Research in Mathematics Education*, *14*, 107-121.
- Christou, K. P. (2015b). Natural number bias in operations with missing numbers. *ZDM Mathematics Education*, *47*(5), 747-758. <https://doi.org/10.1007/s11858-015-0675-6>
- Christou, K. P. (2017). The natural number bias in arithmetic operations – Beyond the size of the numbers. In B. Cvetkova Dimov (Ed.), *3rd International Conference "Education Across Borders – Education and Research Across Time and Space"* (pp. 546-553). Bitola, Northern Macedonia: University "St. Kliment Ohridski".
- Christou, K. P., & Vosniadou, S. (2012). What kinds of numbers do students assign to literal symbols? Aspects of the transition from arithmetic to algebra. *Mathematical Thinking and Learning*, *14*(1), 1-27. <https://doi.org/10.1080/10986065.2012.625074>
- Christou, K. P., Vosniadou, S., & Vamvakoussi, X. (2007). Students' interpretations of literal symbols in algebra. In S. Vosniadou, A. Baltas, & X. Vamvakoussi (Eds.), *Reframing the conceptual change approach in learning and instruction* (pp. 283-297). Amsterdam, The Netherlands: Elsevier.
- Depaepe, F., Torbeyns, J., Vermeersch, N., Janssens, D., Janssen, R., Kelchtermans, G., . . . Van Dooren, W. (2015). Teachers' content and pedagogical content knowledge on rational numbers: A comparison of prospective elementary and lower secondary school teachers. *Teaching and Teacher Education*, *47*, 82-92. <https://doi.org/10.1016/j.tate.2014.12.009>
- Depaepe, F., Van Roy, P., Torbeyns, J., Kleickmann, T., Van Dooren, W., & Verschaffel, L. (2018). Stimulating pre-service teachers' content and pedagogical content knowledge on rational numbers. *Educational Studies in Mathematics*, *99*(2), 197-216. <https://doi.org/10.1007/s10649-018-9822-7>
- Desmet, L., Grégoire, J., & Mussolin, C. (2010). Developmental changes in the comparison of decimal fractions. *Learning and Instruction*, *20*, 521-532. <https://doi.org/10.1016/j.learninstruc.2009.07.004>
- DeWolf, M., & Vosniadou, S. (2011). The whole number bias in fraction magnitude comparisons with adults. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*. Retrieved from <https://escholarship.org/uc/item/0vw4p1c2>
- Dimitrakopoulou, S. A., & Christou, K. P. (2018). Treatment of variables in Greek mathematics textbooks and the different ways that students interpret them. *Journal for Research in Mathematics Education / Erevna gia ti didaktiki ton mathimatikon*, *11*, 31-52. <https://doi.org/10.12681/enedim.18938>
- Dixon, J. A., Deets, J. K., & Bangert, A. (2001). The representations of the arithmetic operations include functional relationships. *Memory & Cognition*, *29*(3), 462-477. <https://doi.org/10.3758/BF03196397>
- Fischbein, E. (1987). *Intuition in science and mathematics*. Dordrecht, The Netherlands: Reidel.
- Fischbein, E., Deri, M., Nello, M., & Marino, M. (1985). The role of implicit models in solving verbal problems in multiplication and division. *Journal for Research in Mathematics Education*, *16*, 3-17. <https://doi.org/10.2307/748969>

- Gelman, R. (2000). The epigenesis of mathematical thinking. *Journal of Applied Developmental Psychology, 21*, 27-37. [https://doi.org/10.1016/S0193-3973\(99\)00048-9](https://doi.org/10.1016/S0193-3973(99)00048-9)
- Gillard, E., Van Dooren, W., Schaeken, W., & Verschaffel, L. (2009). Dual processes in the psychology of mathematics education and cognitive psychology. *Human Development, 52*, 95-108. <https://doi.org/10.1159/000202728>
- Gómez, D. M., Jiménez, A., Bobadilla, R., Reyes, C., & Dartnell, P. (2014). Exploring fraction comparison in school children. In S. Oesterle, P. Liljedahl, C. Nicol, & D. Allan (Eds.), *Proceedings of the Joint Meeting 3 - 185 of PME 38 and PME-NA 36* (Vol. 3, pp. 185-192). Vancouver, Canada: PME.
- Graeber, A. O., Tirosh, D., & Glover, R. (1989). Preservice teachers' misconceptions in solving verbal problems in multiplication and division. *Journal for Research in Mathematics Education, 20*, 95-102. <https://doi.org/10.2307/749100>
- Greer, B. (1987). Nonconservation of multiplication and division involving decimals. *Journal for Research in Mathematics Education, 18*(1), 37-45. <https://doi.org/10.2307/749535>
- Greer, B. (1994). Extending the meaning of multiplication and division. In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 61-88). New York, NY, USA: State University of New York Press.
- Harel, G., & Confrey, J. (1994). *The development of multiplicative reasoning in the learning of mathematics*. New York, NY, USA: State University of New York Press.
- Hart, K. M. (1981). *Children's understanding of mathematics* (pp. 11-16). London, United Kingdom: John Murray.
- Hartnett, P., & Gelman, R. (1998). Early understandings of numbers: Paths or barriers to the construction of new understandings? *Learning and Instruction, 8*(4), 341-374. [https://doi.org/10.1016/S0959-4752\(97\)00026-1](https://doi.org/10.1016/S0959-4752(97)00026-1)
- Iuculano, T., & Butterworth, B. (2011). Rapid communication: Understanding the real value of fractions and decimals. *Quarterly Journal of Experimental Psychology, 64*(11), 2088-2098. <https://doi.org/10.1080/17470218.2011.604785>
- Izsák, A., & Beckmann, S. (2018). Using equations to develop a coherent approach to multiplication and measurement. In E. Bergqvist, M. Österholm, C. Granberg, & L. Sumpter (Eds.), *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 155-162). Umeå, Sweden: PME.
- Krueger, L. E. (1986). Why  $2 \times 2 = 5$  looks so wrong: On the odd-even rule in product verification. *Memory & Cognition, 14*(2), 141-149. <https://doi.org/10.3758/BF03198374>
- Krueger, L. E., & Hallford, E. W. (1984). Why  $2 + 2 = 5$  looks so wrong: On the odd-even rule in sum verification. *Memory & Cognition, 12*(2), 171-180. <https://doi.org/10.3758/BF03198431>
- Leron, U., & Hazzan, O. (2006). The rationality debate: Application of cognitive psychology to mathematics education. *The Mathematics Educator, 62*, 105-126. <https://doi.org/10.1007/s10649-006-4833-1>
- Mazzocco, M. M., & Devlin, K. T. (2008). Parts and 'holes': Gaps in rational number sense among children with vs. without mathematical learning disabilities. *Developmental Science, 11*(5), 681-691. <https://doi.org/10.1111/j.1467-7687.2008.00717.x>
- Merenluoto, K., & Lehtinen, E. (2004). Number concept and conceptual change: Towards a systemic model of the processes of change. *Learning and Instruction, 14*, 519-534. <https://doi.org/10.1016/j.learninstruc.2004.06.016>

- Moss, J. (2005). Pipes, tubes, and beakers: New approaches to teaching the rational-number system. In S. Donovan & J. D. Bransford (Eds.), *How students learn: History, mathematics, and science in the classroom* (pp. 309-349). Washington, DC, USA: National Academy Press.
- Moutier, S., & Houdé, O. (2003). Judgement under uncertainty and conjunction fallacy inhibition training. *Thinking & Reasoning*, 9(3), 185-201. <https://doi.org/10.1080/13546780343000213>
- Nesher, P., & Peled, I. (1986). Shifts in reasoning. *Educational Studies in Mathematics*, 17(1), 67-79. <https://doi.org/10.1007/BF00302379>
- Ni, Y. J., & Zhou, Y.-D. (2005). Teaching and learning fraction and rational numbers: The origins and implications of whole number bias. *Educational Psychologist*, 40(1), 27-52. [https://doi.org/10.1207/s15326985ep4001\\_3](https://doi.org/10.1207/s15326985ep4001_3)
- Obersteiner, A., Van Hoof, J., Verschaffel, L., & Van Dooren, W. (2016). Who can escape the natural number bias in rational number tasks? A study involving students and experts. *British Journal of Psychology*, 107(3), 537-555. <https://doi.org/10.1111/bjop.12161>
- Onslow, B. (1990). Overcoming conceptual obstacles: The qualified use of a game. *School Science and Mathematics*, 90(7), 581-592. <https://doi.org/10.1111/j.1949-8594.1990.tb12032.x>
- Prather, R. W., & Alibali, M. W. (2008). Understanding and using principles of arithmetic: Operations involving negative numbers. *Cognitive Science*, 32(2), 445-457. <https://doi.org/10.1080/03640210701864147>
- Prediger, S. (2008). The relevance of didactic categories for analysing obstacles in conceptual change: Revisiting the case of multiplication of fractions. *Learning and Instruction*, 18(1), 3-17. <https://doi.org/10.1016/j.learninstruc.2006.08.001>
- Rees, R., & Barr, G. (1984). *Diagnosis and prescription: Some common math problems*. London, United Kingdom: Harper & Row.
- Resnick, L. B., Nesher, P., Leonard, F., Magone, M., Omanson, S., & Peled, I. (1989). Conceptual bases of arithmetic errors: The case of decimal fractions. *Journal for Research in Mathematics Education*, 20, 8-27. <https://doi.org/10.2307/749095>
- Roell, M., Viarouge, A., Houdé, O., & Borst, G. (2017). Inhibitory control and decimal number comparison in school-aged children. *PLOS ONE*, 12(11), Article e0188276. <https://doi.org/10.1371/journal.pone.0188276>
- Roell, M., Viarouge, A., Houdé, O., & Borst, G. (2019). Inhibition of the whole number bias in decimal number comparison: A developmental negative priming study. *Journal of Experimental Child Psychology*, 177, 240-247. <https://doi.org/10.1016/j.jecp.2018.08.010>
- Saxe, G. B. (2015). *Culture and cognitive development: Studies in mathematical understanding*. New York, NY, USA: Psychology Press.
- Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, 124(2), 209-215. <https://doi.org/10.1016/j.cognition.2012.04.005>
- Siegler, R. S., Fazio, L. K., Bailey, D. H., & Zhou, X. (2013). Fractions: The new frontier for theories of numerical development. *Trends in Cognitive Sciences*, 17(1), 13-19. <https://doi.org/10.1016/j.tics.2012.11.004>

- Smith, C. L., Solomon, G. E. A., & Carey, S. (2005). Never getting to zero: Elementary school students' understanding of the infinite divisibility of number and matter. *Cognitive Psychology*, *51*, 101-140. <https://doi.org/10.1016/j.cogpsych.2005.03.001>
- Stazyk, E. H., Ashcraft, M. H., & Hamann, M. S. (1982). A network approach to mental multiplication. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(4), 320-335. <https://doi.org/10.1037/0278-7393.8.4.320>
- Thorndike, E. L. (1922). *The psychology of arithmetic*. New York, NY, USA: Macmillan.
- Vamvakoussi, X., Christou, K. P., Mertens, L., & Van Dooren, W. (2011). What fills the gap between discrete and dense? Greek and Flemish students' understanding of density. *Learning and Instruction*, *21*(5), 676-685. <https://doi.org/10.1016/j.learninstruc.2011.03.005>
- Vamvakoussi, X., Christou, K. P., & Vosniadou, S. (2018). Bridging psychological and educational research on rational number knowledge. *Journal of Numerical Cognition*, *4*(1), 84-106. <https://doi.org/10.5964/jnc.v4i1.82>
- Vamvakoussi, X., Van Dooren, W., & Verschaffel, L. (2012). Naturally biased? In search for reaction time evidence for a natural number bias in adults. *The Journal of Mathematical Behavior*, *31*, 344-355. <https://doi.org/10.1016/j.jmathb.2012.02.001>
- Vamvakoussi, X., Van Dooren, W., & Verschaffel, L. (2013). Educated adults are still affected by intuitions about the effect of arithmetical operations: Evidence from a reaction-time study. *Educational Studies in Mathematics*, *82*(2), 323-330. <https://doi.org/10.1007/s10649-012-9432-8>
- Vamvakoussi, X., & Vosniadou, S. (2010). How many decimals are there between two fractions? Aspects of secondary school students' understanding of rational numbers and their notation. *Cognition and Instruction*, *28*(2), 181-209. <https://doi.org/10.1080/07370001003676603>
- Van Dooren, W., & Inglis, M. (2015). Inhibitory control in mathematical thinking, learning and problem solving: A survey. *ZDM*, *47*(5), 713-721. <https://doi.org/10.1007/s11858-015-0715-2>
- Van Dooren, W., Lehtinen, E., & Verschaffel, L. (2015). Unraveling the gap between natural and rational numbers. *Learning and Instruction*, *37*, 1-4. <https://doi.org/10.1016/j.learninstruc.2015.01.001>
- Van Hoof, J., Vandewalle, J., Verschaffel, L., & Van Dooren, W. (2015). In search for the natural number bias in secondary school students' interpretation of the effect of arithmetical operations. *Learning and Instruction*, *37*, 30-38. <https://doi.org/10.1016/j.learninstruc.2014.03.004>
- Vosniadou, S., Vamvakoussi, X., & Skopeliti, E. (2008). The framework theory approach to conceptual change. In S. Vosniadou (Ed.), *Handbook of research on conceptual change* (pp. 3-34). Mahwah, NJ, USA: Lawrence Erlbaum Associates.
- Winkelman, J. H., & Schmidt, J. (1974). Associative confusions in mental arithmetic. *Journal of Experimental Psychology*, *102*(4), 734-736. <https://doi.org/10.1037/h0036103>

### Appendix

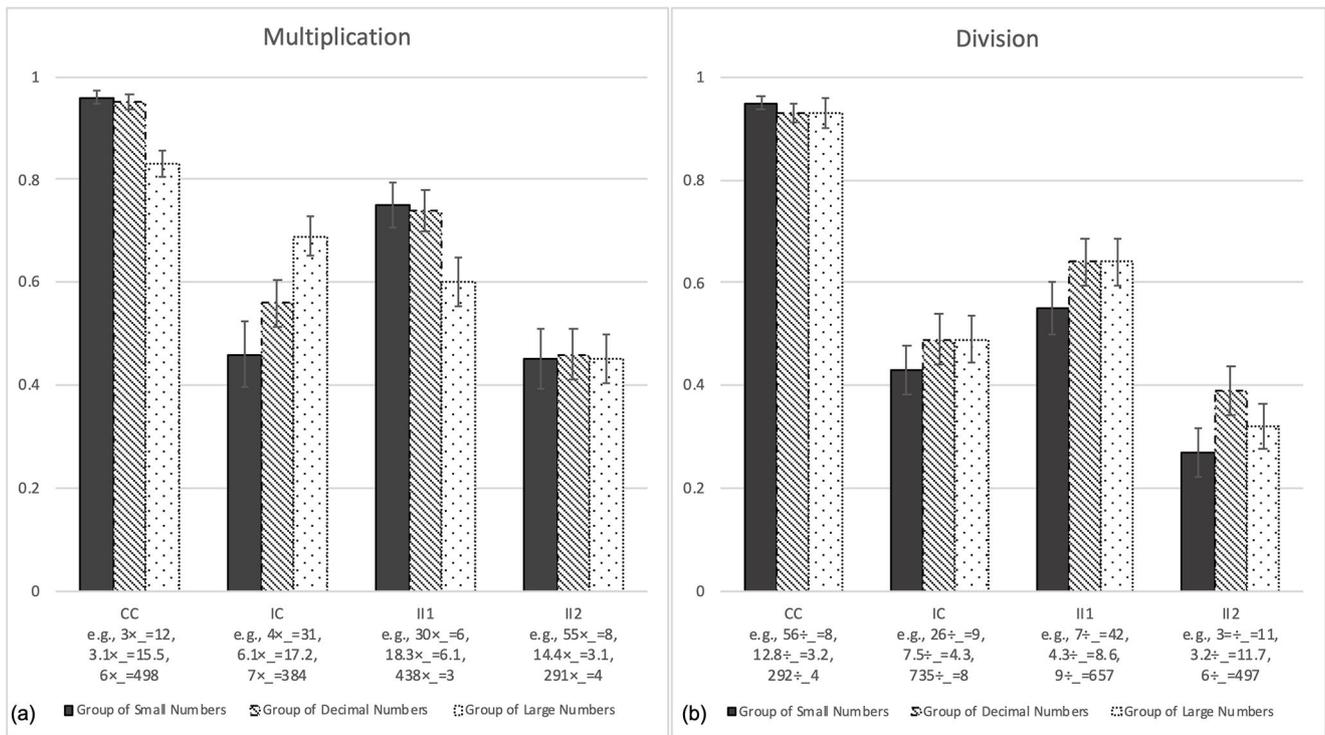


Figure A.1. Accuracy rates by item type and group for (a) multiplication and (b) division problems.

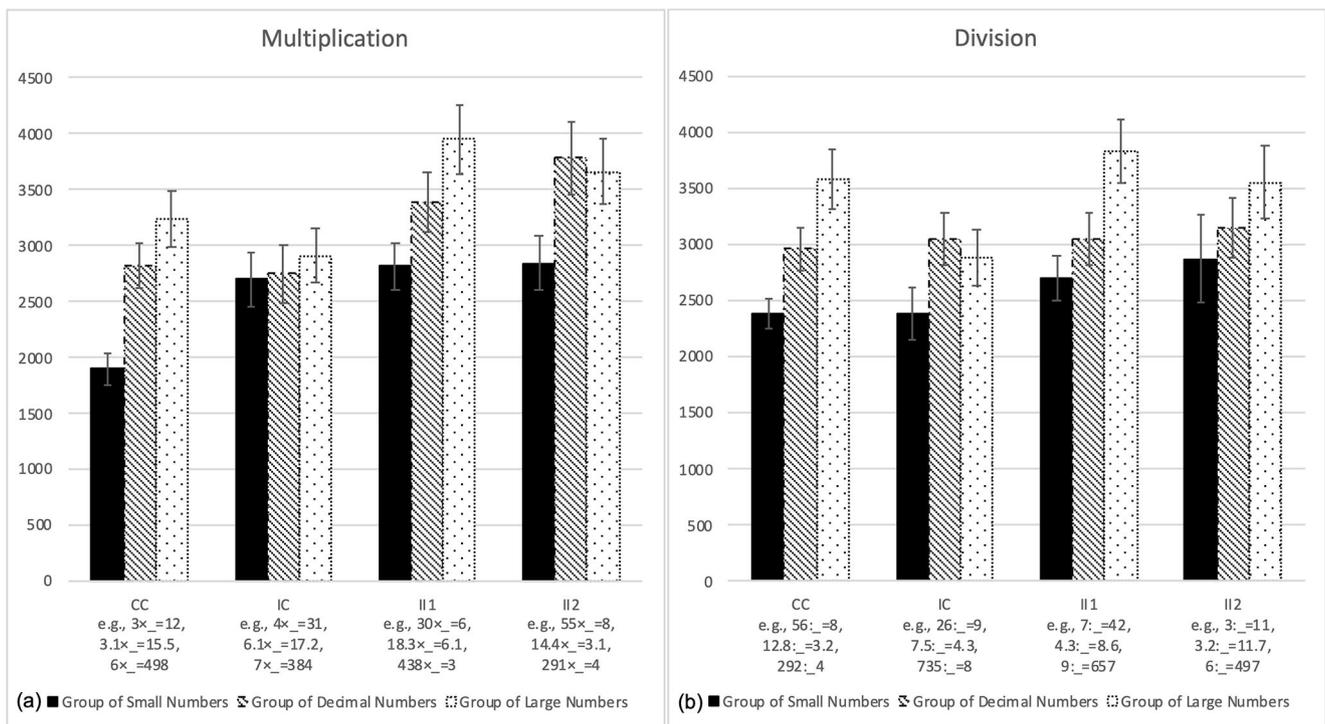


Figure A.2. Reaction time by item type and group for (a) multiplication and (b) division problems.