

# Arithmetic word problem solving: evidence for a magnitude-based mental representation

Josetxu Orrantía · David Múñez

Published online: 17 August 2012  
© Psychonomic Society, Inc. 2012

**Abstract** Previous findings have suggested that number processing involves a mental representation of numerical magnitude. Other research has shown that sensory experiences are part and parcel of the mental representation (or “simulation”) that individuals construct during reading. We aimed at exploring whether arithmetic word-problem solving entails the construction of a mental simulation based on a representation of numerical magnitude. Participants were required to solve word problems and to perform an intermediate figure discrimination task that matched or mismatched, in terms of magnitude comparison, the mental representations that individuals constructed during problem solving. Our results showed that participants were faster in the discrimination task and performed better in the solving task when the figures matched the mental representations. These findings provide evidence that an analog magnitude-based mental representation is routinely activated during word-problem solving, and they add to a growing body of literature that emphasizes the experiential view of language comprehension.

**Keywords** Mental representation · Word problem solving · Magnitude representation · Number representation · Grounded cognition

Psycholinguistic theories have assumed that understanding a text consists not just in the construction of its propositional structure (e.g., Kintsch & van Dijk, 1978), but is only fully

understood if a reader has constructed a mental representation or “simulation” of the situation that the text describes. This view of understanding as mental simulation is rooted in experiential theories of language comprehension, according to which the mental representations involved in comprehension may take the form of perceptual symbols, which are directly derived from perceptual experiences (Barsalou, 1999; Glenberg & Robertson, 1999; Zwaan, 1999). Nevertheless, until now, no study has directly addressed the relevance of this view in the domain of arithmetic word-problem solving. The present study constitutes a first step to fill this gap.

Arithmetic word problems verbally describe a situation (Verschaffel, Greer, & De Corte, 2000). Therefore, this type of text can be considered, like any other text, as a valid discourse entity, since not only arithmetic skills are involved in solving these problems, but also understanding of both the language involved and the situation denoted by its verbal description (Kintsch, 1988). The traditional theoretical position postulates that the representation of the text of the problem is built through the transformation of linguistic inputs into conceptual representations of their meaning in the form of propositional symbols—an abstract symbolic structure (e.g., Kintsch & Greeno, 1985). In this regard, Thevenot (2010) has recently provided evidence that the mental representation built by individuals to solve arithmetic word problems preserves only the relationship between the elements of the situation described in the wording. However, whether the nature of this mental representation goes beyond an abstract propositional structure remains unknown. In this sense, the experiential view of language comprehension is particularly suitable for accounting for the nature of the mental representation constructed during word-problem solving. As is proposed in this theoretical perspective, individuals understand language by constructing a perceptual simulation of the objects and events described. This mental representation is based on perceptual symbols that

---

J. Orrantía (✉) · D. Múñez  
Department of Educational and Developmental Psychology,  
Universidad de Salamanca,  
Paseo de Canalejas 169,  
37008 Salamanca, Spain  
e-mail: orrantia@usal.es

D. Múñez  
e-mail: davidm@usal.es

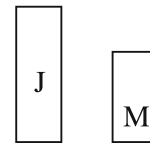
arise from analog perceptual experiences, which are activated while words and sentences are processed. Thus, considerable experimental support has provided evidence that readers mentally represent motion, orientation, or shape information during reading (for an overview, see Zwaan, 2004), which confirms that language comprehension strongly depends on the reader's experience with, and knowledge of, the world. On the basis of this theoretical stance, we suggest that the mental representation constructed during word-problem solving might not be an abstract symbolic structure, but rather an analog representation (or "simulation") that involves a reenactment of solvers' experience with the processing of magnitude information for quantities.

Thus, much like orientation or shape, numerical magnitude is a basic feature of the environment to which individuals appear to attend (Piazza, 2010), and its representation appears to emerge early in human development (Xu & Spelke, 2000). This awareness of numerical magnitude is thought to serve as a foundation on which symbolic numerical thinking is built. Evidence has suggested that the development of the ability to represent and process numerical symbols (such as number words and Arabic numerals) is grounded on these preexisting magnitude representations (see Ansari, 2008, for a review). In the course of learning, when children are first being introduced to formal arithmetic, mental representations of numerosity are strongly related to physical quantity representations. Children construct external representations using objects or fingers to count sets, to manipulate these sets (by combining, adding, comparing, . . .), or to calculate. Thus, individuals can accumulate a lot of experiences with these external representations on which the numerical symbols are mapped, so that these experiences, which are closely bound to perception and action, might leave experiential traces that could be reactivated during problem processing.

Although the role of magnitude information in word-problem processing has attracted little systematic research, many studies have demonstrated that number processing involves a mental representation of numerical magnitude. A widely replicated effect in this regard is the "numerical distance effect" (Moyer & Landauer, 1967), which indicates that it is more time-consuming to compare two numbers when the numerical distance between them is small (e.g., 7 vs. 9) than when it is large (e.g., 3 vs. 8). This effect has also been reported in other, nonnumerical magnitude dimensions, such as line length or geometrical shapes (for a review, see Cohen Kadosh, Lammertyn, & Izard, 2008), indicating that the magnitude representation of numbers is shared with other nonnumerical magnitude dimensions (Walsh, 2003). It has also been demonstrated that numerical and nonnumerical quantities interact with each other. In the numerical Stroop task, participants take more time to

judge either numerical or physical size when the number magnitude and the physical size of a digit are incongruent (Schwarz & Ischebeck, 2003; see also Gebuis, Cohen Kadosh, de Haan, & Henick, 2009, for a nonsymbolic size congruity task). Thus, the incompatible magnitude of one task-irrelevant dimension would interfere with the judgment on the other magnitude dimension. In addition to behavioral research, a number of neuroimaging studies on number processing have demonstrated that the intraparietal sulcus (IPS) is involved in processing both numerical and nonnumerical magnitudes (see Brannon, 2006, and Cohen Kadosh et al., 2008, for reviews). Thus, IPS activations have been observed with magnitudes presented as symbolic stimuli (i.e., as Arabic or verbal numerals), nonsymbolic discrete stimuli (e.g., dot patterns), or continuous stimuli (e.g., length of lines).

Despite the fact that most studies on quantity processing have focused on simple tasks (e.g., magnitude comparison), it would be expected that in higher cognitive tasks, such as word-problem solving, a representation of numerical magnitude related to the situation described would also be activated. Given that in a problem such as "John has 5 marbles more than Mary; If John has 8 marbles, how many marbles does Mary have?" the first sentence leads to a mental representation that preserves the relationship between the terms, it might be possible for experiential traces, related to the processing of numerical magnitude, to be reactivated to produce a mental simulation of the described situation, similar to:



This mental representation would be feasible, since solvers likely engage in magnitude comparison to verify which protagonist possesses more target objects, and in addition, converging evidence has indicated that a shared magnitude representation underlies different quantity dimensions. In fact, Lee et al. (2007) found common activation of the IPS associated with symbolic and non-symbolic representations during reading of a word problem. Specifically, participants were asked to transform information from a word problem (e.g., James has 50 fewer watches than Mike) to either a symbolic equation (e.g.,  $J = M - 50$ ) or a nonsymbolic representation based on a diagram made up of rectangles to represent relationships in the word problem (e.g., a rectangle representing Mike, and another, shorter one representing James). The authors found activation in the IPS for both the symbolic and nonsymbolic conditions, which suggests that there is a common mechanism for magnitudes even in the context of word-problem solving.

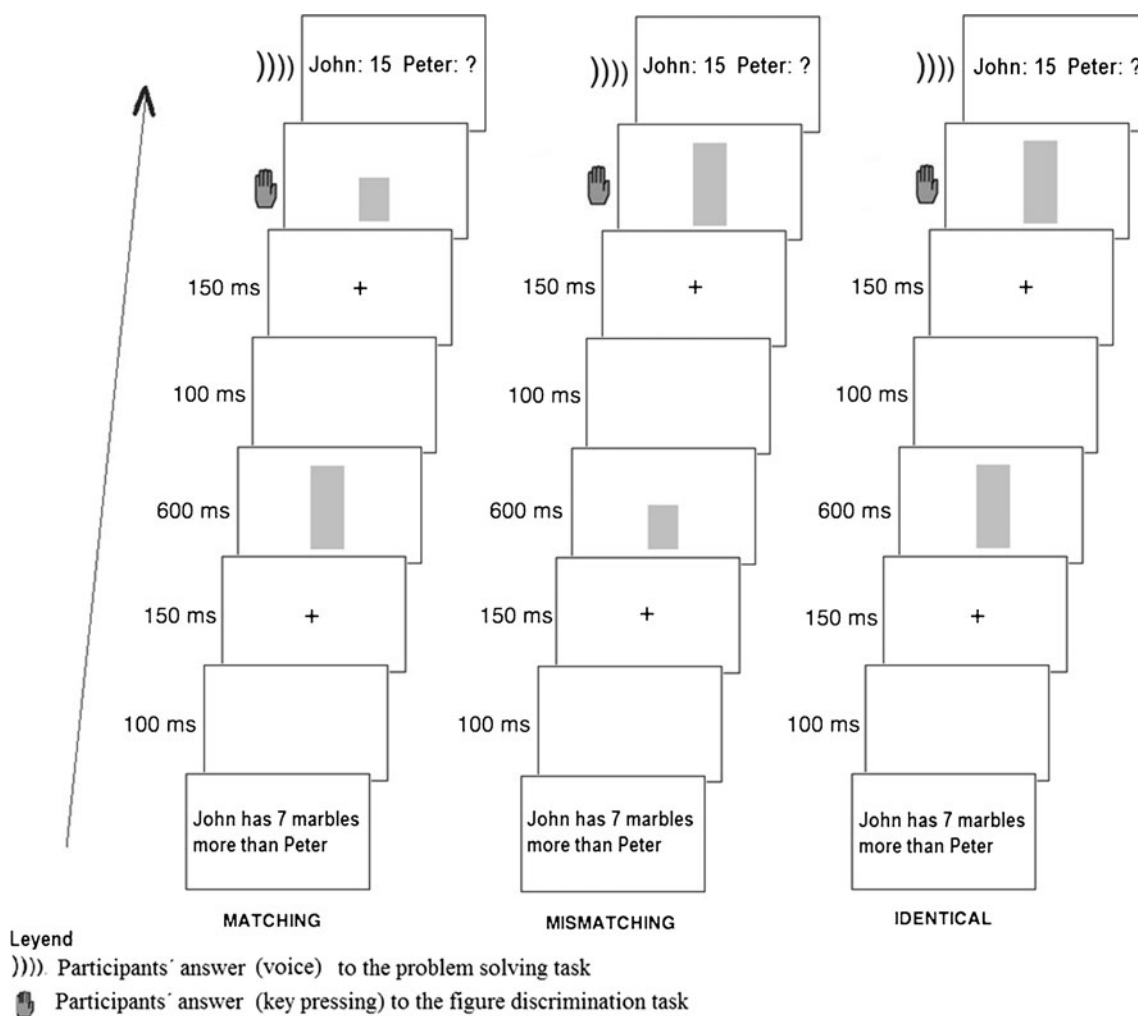
## The present study

The purpose of this study was to assess whether a magnitude-based mental representation is routinely activated when arithmetic word problems are solved. This representation would go beyond a conceptual representation in the form of propositional symbols, to reflect in terms of magnitudes the situation that the word problem describes. Participants were required to solve “compare” word problems (see Riley, Greeno, & Heller, 1983) and to perform an intermediate discrimination task not related to the solving process (see Fig. 1). On each trial, participants read the first sentence of the problem, which stated the relationship between two variables, then judged whether two sequentially presented figures were perceptually the same or different, and finally read the remaining information of the problem in order to solve it. Importantly, by using a same–different task instead of a magnitude comparison task (i.e., to decide whether the second figure was larger or smaller than the first), any explicit cue related to magnitude would be eliminated. Therefore, the task would reflect the construction

of an automatic mental representation from the target sentence (Orrantia, Rodríguez, & Vicente, 2010; Tzelgov & Ganor-Stern, 2005), since the figures’ discrimination would not be part of the task requirements (i.e., problem solving).

Assuming that the first rectangle would be mapped onto the first person named in the word problem and the second rectangle mapped onto the second person named in the problem (see the [General Discussion](#)), the sequences of figures would either match or mismatch the possible mental representation constructed from the relational sentence of the problem. In the matching condition, the figures depicted the relationship between the variables contained in the relational sentence (John > Peter in Fig. 1), whereas in the mismatching condition, the figures depicted the opposite relationship (John < Peter). In keeping with this, the identical condition (John = Peter) was also considered as a mismatching condition.

In addition, the compare problems were either consistent (CL) or inconsistent (IL) language problems (Lewis & Mayer, 1987). In a CL problem (e.g., “John has 7 marbles



**Fig. 1** Example of the three types of experimental trials in Experiment 2

more than Peter; John: ? Peter: 8”), the unknown variable (i.e., John) is the grammatical subject of the relational sentence, and the relational term (i.e., “more than”) is consistent with the required arithmetic operation (i.e., addition). In an IL problem (e.g., “John has 7 marbles more than Peter; John: 15 Peter: ?”), the unknown variable (i.e., Peter) is the object of the relational sentence, and the relational term (i.e., “more than”) is in conflict with the required arithmetic operation (i.e., subtraction). A number of empirical studies have shown that IL problems are considerably more difficult to solve than CL problems (e.g., Hegarty, Mayer, & Monk, 1995; Verschaffel, 1994; Verschaffel, De Corte, & Pauwels, 1992), possibly because, according to Lewis and Mayer’s (1987) “consistency hypothesis,” solvers prefer the order included in CL problems, in which the unknown variable is the subject of the relational sentence. Thus, when given an IL problem, solvers are assumed to mentally rearrange the relational sentence by reversing the subject and object as well as the relational term. This mental rearrangement procedure is a cognitively demanding one, which will result in greater difficulty for IL than for CL problems.

If solvers activate a magnitude-based representation, the discrimination task on pairs of figures matching the mental representation created from the relational sentence would be faster than on pairs mismatching the mental representation. Furthermore, match/mismatch effects would be obtained for the problem-solving task. Since the relational sentence was not displayed to participants at the time of problem solving, the matching condition may act as an (implicitly presented) external representation reinforcing the salience of the mental representation created from the relational sentence. Thus, problem solving would be faster and less error-prone in the matching than in the mismatching conditions. Nevertheless, this effect would be larger for IL than for CL problems, since IL problems require additional processing to rearrange the relational sentence. In this sense, the figures in the matching condition could serve as a type of external memory that would reduce the demands on working memory in this rearrangement procedure.

Four experiments were designed to examine these predictions. The aims of Experiments 1a and 1b were to establish a baseline in order to interpret the results of discrimination and problem-solving tasks in terms of facilitation and/or interference due to the matching/mismatching condition. To do this, in Experiment 1a participants performed a discrimination task on pairs of figures following presentation of a neutral sentence—that is, one that did not activate magnitudes. In Experiment 1b, participants solved CL and IL compare word problems with no intermediate discrimination task. Experiment 2 was designed to provide evidence that an analog magnitude-based mental representation is activated while solving arithmetic word problems. Participants were required to solve compare word problems

and to perform an intermediate figure discrimination task that matched or mismatched, in terms of magnitude, the mental representation of the problem (see Fig. 1). Finally, an additional control study (Exp. 3) tested whether the effects of facilitation/interference on problem solving were due to the nature of the mental representation induced by the discrimination task. For this purpose, participants solved compare word problems in which the sequence of figures was replaced by a sequence of colors.

## Preliminary Experiments 1a and 1b

### Experiment 1a

#### Method

*Participants* A group of 32 psychology undergraduates took part in this experiment.

*Materials* A total of 48 experimental sentences, each followed by a figure discrimination task and probe-word recognition, were used in the experiment. All of the sentences (which varied in length from six to eight words) mentioned one or two differently named characters engaged in an action. The sentences did not (explicitly or implicitly) involve any reference to quantity. The probe word could be any word from the sentence. Following presentation of the sentence was the figure discrimination task. Two types of sequences of figures were presented: nonidentical and identical. As figures, we used ten monochromatic gray rectangles that were 820 pixels wide. The heights of the different rectangles increased by intervals of 100 pixels, ranging from 480 to 1,480 pixels. To avoid differences in visual scanning speed, the figures’ lower edges were set on screen at the same level ( $y = -150$ ), and nonidentical pairs differed by 300 pixels in height, which was enough to discriminate the two figures as nonidentical. Each sequence consisted of one pair of figures selected from 22 possible pairings (10 identical and 12 nonidentical, which excluded the two pairs in which the smallest and largest figures were presented first in the order). This pairing procedure aimed at preventing participants from anticipating the size of the second figure from the size of the first one. After the discrimination task, the probe word appeared. The trials were grouped into an initial warm-up block of five trials, followed by one experimental block of 48 trials. The experimental block contained equal numbers of identical and nonidentical sequences and probe-present and probe-absent trials.

*Procedure* The stimuli were presented using SuperLab software, which ran on computers with 15-in. monitors (this same apparatus was used in all experiments). On each trial, the sentence (Geneva 36-point font) was presented in the

middle of the screen. Pressing the space bar caused the sentence to be erased and the discrimination task to be presented. Following a blank screen (100 ms) and a fixation cross (150 ms), a sequence of two monochromatic gray rectangles was shown. The first rectangle was shown for 600 ms in the middle of the screen. After another blank screen (100 ms) and fixation cross (150 ms), the second rectangle was presented in the same place as the first one. Participants indicated via keypresses whether the two figures were the same or different (the response-key assignment was counterbalanced between subjects). Then, a single probe word appeared in the middle of the screen. Participants were required to indicate as quickly as possible whether the probe word had occurred in the preceding sentence by pressing the key that was also assigned to a SAME response in the figure discrimination task, if it had, and the key that was assigned to a DIFFERENT response, if it had not. After an information screen (“press the space bar to continue”) and a blank screen (2, 500 ms), the next trial started.

### Results

Response times that were 2.5 standard deviations from the mean or those following incorrect responses in the discrimination task were discarded. This eliminated no more than 2 % of the data. No significant differences were found between the responses to identical and nonidentical figures (597 and 608 ms, respectively;  $p > .15$ ).

### Experiment 1b

#### Method

**Participants** A group of 25 psychology undergraduates participated in this experiment.

**Materials** The participants were presented with 48 experimental arithmetic word problems. The problems belonged to the Compare 3 to Compare 6 categories (described below) according to the classification scheme of Riley et al. (1983), and they differed in their situations, the names of the protagonists, and the nature of the objects. Whereas in Compare 3 (including the term “more than”) and Compare 4 (including the term “less than”) the unknown variable is the grammatical subject of the relational sentence, in Compare 5 (including the term “more than”) and Compare 6 (including the term “less than”), the unknown variable is the object of the relational sentence. Therefore, Compare 3 and 4 correspond to CL problems, and Compare 5 and 6 to IL problems. The first sentence of each problem was the relational sentence (e.g., “John has 7 marbles more than Peter”). After presentation of the relational sentence, the remaining information of the problem appeared in a line that involved the characters and

the data (e.g., “John: 15 Peter: ?”; see Fig. 1). We used this format because in a standard format (e.g., “if John has 8 marbles, how many marbles does Peter have?”) the self-presentation time for the question contains the resolution time in addition to the reading time; the question could even be inferred, so that its reading time would be much faster. All of this would result in a distortion of the response times. In order to reduce additional delay between the relational sentence and the answer, the order of presentation of the characters in this segment of the problem was the same as in the relational sentence. Since the unknown variable could be the grammatical subject or object of the relational sentence, half of the problems were CL, and the other half IL problems.

The operands used in the word problems consisted of pairs of numbers selected from 36 possible pairings of the digits 2 through 9 when operand order was ignored. Ties (e.g.,  $3 + 3$ ) and nontie pairs with a sum or minuend of 11 were excluded. To control the problem size effect,<sup>1</sup> operand pairs with a sum or minuend  $\leq 10$  were classified as small, whereas operand pairs with a sum or minuend  $\geq 12$  were classified as large. Thus, half of the word problems contained operands classified as small, and the other half had operands classified as large. Twenty-four filler word problems were included to vary the form of the problems. These filler word problems belonged to the Combine 2 category of Riley et al.’s (1983) classification scheme (e.g., “Joe and Tom have 8 marbles altogether; Joe: 3 Tom: ?”).

**Procedure** On each trial, the relational sentence (Geneva 36-point font) was presented in the middle of the screen. Pressing the space bar caused the relational sentence to be erased and the remaining information of the problem to appear in the middle of the screen. The participants were told to solve the word problem correctly as quickly as possible. A microphone connected to a voice-activated relay and interfaced with the computer registered the latency of the responses. After an information screen (“press the space bar to continue”) and a blank screen (2,500 ms), the next trial started. To familiarize participants with the procedure, they performed four practice trials.

**Scoring** For each participant, two scores were determined for each experimental word problem: (a) the proportion of operation errors—that is, reversal of the numerical operation of the problem (i.e., addition for subtraction or subtraction for addition)—and (b) the response time between the end of

<sup>1</sup> One of the most extensively researched phenomena of cognitive arithmetic is the problem size effect, which refers to the observation that the difficulty of simple arithmetic combinations generally increases with numerical size, and that problems with larger operands (e.g.,  $7 + 8$ ) are solved more slowly than those with smaller operands (e.g.,  $3 + 2$ ; see Ashcraft, 1992, for a review).

reading of the relational sentence and the moment that the answer was given.

### Results and discussion

Mean response times and errors rates were analyzed in a 2 (size: small vs. large operands)  $\times$  2 (consistency: CL vs. IL) repeated measures ANOVA.

**Errors** The analysis indicated a main effect of consistency [ $F(1, 24) = 27.02, MSE = .015, p.0001, \eta_p^2 = .52$ ]; participants made more errors on IL problems than on CL problems (.15 vs. .03, respectively). We found no significant size effect ( $p > .2$ ), possibly because only operation errors were considered.

**Response times** A total of 4.1 % of the trials were excluded as outliers deviating more than 2.5 standard deviations from the mean or due to voice-key errors, with no significant differences between conditions. A main effect of size emerged [ $F(1, 24) = 90.33, MSE = 133, 184, p < .0001, \eta_p^2 = .79$ ]; word problems with small operands (1,822 ms) were solved faster than those with large operands (2,516 ms). This finding is consistent with previous arithmetic research in which the size effect has been found not only in single-digit tasks (see Ashcraft, 1992), but also in more complex arithmetic tasks, such as word problems (e.g., Orrantia, Rodríguez, Múñez, & Vicente, 2012). In addition, a significant main effect of consistency was found [ $F(1, 24) = 63.17, MSE = 101, 538, p < .0001, \eta_p^2 = .73$ ], in which participants solved CL problems more quickly (1,916 ms) than IL problems (2,423 ms). This replicates the results reported by other studies in which a similar consistency effect was found with word problems in a standard format (e.g., Hegarty et al., 1995; Verschaffel et al., 1992), and it supports the validity of the format used in our experiments.

## Experiment 2

### Method

#### Participants

A group of 25 psychology undergraduates were recruited for this experiment.

#### Materials

All participants were presented with 72 experimental arithmetic word problems. The problems belonged to the Compare 3 through 6 categories. The first sentence of each problem held the relational term that activated the mental

representation of the situation described in the problem. Following this sentence was the figure discrimination task. Three types of sequences of figures were presented: non-identical matching mental representations, non identical mismatching mental representations, and identical mismatching mental representations (see Fig. 1). We used the same figures as in Experiment 1a. After the discrimination task, the remaining information about the problem appeared in a line that involved the characters and the data, as in Experiment 1b. Half of the problems were CL, and the other half were IL.

The operands used in the word problems were similar to those used in Experiment 1b. Forty-eight filler word problems were included, to vary the form of the problems and to equalize the numbers of SAME and DIFFERENT responses in the discrimination task. These filler word problems belonged to the Combine 2 category of Riley et al.'s (1983) classification scheme.

### Procedure

On each trial (see Fig. 1), the relational sentence was presented in the middle of the screen. Pressing the space bar caused the sentence to be erased and the discrimination task to be presented. The figure discrimination task was presented with the same time parameters used in Experiment 1a, and participants indicated via keypresses whether the two figures were the same or different. Then, the remaining information of the problem appeared in the middle of the screen. The participants were told to solve the word problem correctly as quickly as possible. A microphone that was connected to a voice-activated relay and interfaced with the computer registered the latencies of the responses. After an information screen ("press the space bar to continue") and a blank screen (2,500 ms), the next trial started. To familiarize participants with the procedure, they performed four practice trials.

### Design and scoring

For the discrimination task, we used a repeated measures design with figures (nonidentical matching vs. nonidentical mismatching vs. identical mismatching) as the only independent variable. Two scores were determined for this task: error rate and response time between the onset of the second rectangle and the moment that the participant registered an answer. For the problem-solving task, the design was a 3 (figures: nonidentical matching vs. nonidentical mismatching vs. identical mismatching) by 2 (size: small vs. large operands) by 2 (consistency: CL vs. IL) repeated measures design. The scores for this task were the proportion of operation errors and the response time between the answer to the discrimination task and the answer to the problem-solving task.

## Results and discussion

### Discrimination task

**Errors** There were 2 % incorrect responses, with no significant differences between experimental conditions ( $F < 1$ ).

**Response times** Response times deviating 2.5 standard deviations from a participant's cell means per condition (1.5 % of trials) were discarded and did not differ between conditions. The analysis showed a significant effect of figures [ $F(2, 48) = 4.84, MSE = 7,966, p < .012, \eta_p^2 = .16$ ]. Planned comparisons indicated that participants discriminated faster in the nonidentical matching condition (615 ms) than in the nonidentical mismatching (645 ms;  $p < .0001$ ) and identical mismatching (646 ms;  $p < .02$ ) conditions. This finding supports our hypothesis that a magnitude-based mental representation is routinely activated during problem solving. The results suggest that the processing of the relational sentence of the problem would activate a magnitude-based representation of the relationship between the terms. This mental representation would elicit priming effects on the discrimination task. These effects could be due to facilitation in the matching condition, interference in the mismatching conditions, or both. However, the results obtained in the preliminary study (Exp. 1a) support the idea that interference effects are more likely to occur than facilitation effects. A combined analysis of the results of Experiments 1a and 2 (with Experiment as a between-subjects factor and Nonidentical and Identical Figures as a within-subjects factor) showed that, when nonidentical matching and identical mismatching conditions were considered in Experiment 2, only the interaction effect was significant ( $p < .005$ ), due to increased response times in the identical mismatching condition; when nonidentical mismatching and identical mismatching conditions were considered in Experiment 2, the interaction effect was not found, but response

times in the discrimination task were higher in Experiment 2 than those in Experiment 1a.

### Problem-solving task

**Errors** The error rates are presented in Table 1. The analysis indicated a main effect of consistency [ $F(1, 24) = 21.34, MSE = .011, p < .0001, \eta_p^2 = .47$ ]: Participants made more errors on IL than on CL problems (.08 vs. .02, respectively). Although the main effect of figures did not reach significance [ $F(2, 48) = 3.01, MSE = .006, p = .059$ ], there was a significant interaction between figures and consistency [ $F(2, 48) = 5.32, MSE = .006, p < .008, \eta_p^2 = .18$ ]. Planned comparisons showed that, whereas no significant differences were found between CL and IL problems in the nonidentical matching condition [ $F(1, 24) = 2.75, p > .11$ ], participants made significantly more errors on IL problems than on CL problems in the nonidentical mismatching [ $F(1, 24) = 28.25, p < .0001, \eta_p^2 = .54$ ] and identical mismatching [ $F(1, 24) = 6.38, p < .01, \eta_p^2 = .21$ ] conditions. In addition, we found an interaction between figures and size [ $F(2, 48) = 5.44, MSE = .006, p < .001, \eta_p^2 = .19$ ]; the effect of figures was significant for word problems with large operands [ $F(2, 23) = 10.43, p < .001, \eta_p^2 = .48$ ], owing to an advantage for the nonidentical matching as compared to the nonidentical mismatching ( $p < .0001$ ) and identical mismatching ( $p < .03$ ) conditions. Word problems with small operands did not present this effect [ $F(2, 23) = 1.65, p > .20$ ].

**Response times** Five percent of the trials were excluded as outliers deviating more than 2.5 standard deviations from the mean or due to voice-key errors, with no significant differences between conditions. The corresponding response time data are presented in Table 1. A main effect of size emerged [ $F(1, 24) = 59.92, MSE = 1,705,541, p < .0001, \eta_p^2 = .70$ ]; word problems with small operands (2,335 ms) were solved faster than those with large operands

**Table 1** Mean error rate and response time (and typical error) in the problem solving task (Exp. 2)

	Error Rate		Response Time (ms)	
	Consistent Problems	Inconsistent Problems	Consistent Problems	Inconsistent Problems
Nonidentical matching				
Small operands	.04 (.015)	.06 (.019)	2,220 (145)	2,518 (220)
Large operands	.01 (.009)	.04 (.015)	3,069 (325)	3,231 (251)
Nonidentical mismatching				
Small operands	.01 (.009)	.06 (.019)	2,047 (153)	2,517 (203)
Large operands	.01 (.009)	.15 (.021)	3,222 (250)	4,087 (361)
Identical mismatching				
Small operands	.02 (.016)	.10 (.030)	1,989 (131)	2,717 (181)
Large operands	.04 (.015)	.08 (.022)	3,139 (232)	4,027 (405)

(3,462 ms). In addition, participants solved CL problems (2,614 ms) more quickly than IL problems (3,183 ms) [ $F(1, 24) = 34.29, MSE = 706,488, p < .0001, \eta_p^2 = .58$ ]. We found a main effect of figures [ $F(2, 48) = 15.33, MSE = 94,633, p < .0001, \eta_p^2 = .39$ ]. As expected, participants solved problems in the nonidentical matching condition more quickly (2,759 ms) than those in the nonidentical mismatching (2,968 ms,  $p < .0001$ ) and identical mismatching (2,978 ms,  $p < .001$ ) conditions. This effect was qualified by a significant interaction between figures and consistency [ $F(2, 48) = 6.18, MSE = 367,320, p < .004, \eta_p^2 = .20$ ]. Pairwise comparisons indicated that although IL problems were solved more slowly than CL problems (2,874 ms and 2,644 ms, respectively), the differences did not reach significance in the nonidentical matching condition [ $F(1, 24) = 4.02, p = .056$ ]. In contrast, participants solved IL problems significantly more slowly than CL problems in the nonidentical mismatching [3,302 ms and 2,634 ms, respectively;  $F(1, 24) = 19.59, p < .0001, \eta_p^2 = .45$ ] and identical mismatching [3,372 ms and 2,564 ms, respectively;  $F(1, 24) = 29.96, p < .0001, \eta_p^2 = .55$ ] conditions. The interaction between figures and size was also significant [ $F(2, 48) = 19.63, MSE = 121,307, p < .0001, \eta_p^2 = .45$ ]. As in the error rate analysis, the effect of figures was significant for word problems with large operands [ $F(2, 23) = 21.74, p < .0001, \eta_p^2 = .65$ ], which shows that response times were faster for nonidentical matching than for nonidentical mismatching ( $p < .0001$ ) and identical mismatching ( $p < .0001$ ). This effect was not observed for word problems with small operands [ $F(2, 23) = 2.47, p > .10$ ].

These results confirm our prediction that match/mismatch effects would be obtained for the problem-solving task. As expected, participants were faster and less error-prone to solve word problems in the matching condition than in the mismatching conditions, and this effect was larger for IL than for CL problems. In other words, the differences between IL and CL problems were smaller in the matching than in the mismatching conditions. Remember that IL problems are more difficult than CL problems because IL problems increase the demands on the solvers' working memory, due to the extra processing required to mentally rearrange the relational sentence. The results of the present experiment suggest that the figures in the matching condition acted as a type of external representation that reduced the working memory demands due to the additional processing involved in solving IL problems. Accordingly, match/mismatch effects were also greater for word problems with large than with small operands, since large operands are more likely to involve arithmetic procedures (e.g., counting, transformations, etc.; see Campbell & Xue, 2001) that are more demanding than direct memory retrieval.

Although the matching condition facilitated the subsequent problem solving by reducing working memory demands, it is also possible that the mismatching conditions interfered with problem solving. The results obtained in Experiment 1b (with no discrimination task) showed a 507-ms advantage for CL problems (1,916 ms) as compared with IL problems (2,423 ms). In the present experiment, the differences between CL and IL problems for the nonidentical matching, nonidentical mismatching, and identical mismatching conditions, respectively, were 230, 668, and 807 ms. A combined analysis of the results of Experiments 1b and 2 showed that, in the matching condition, the differences between CL and IL were significantly reduced in Experiment 2 relative to Experiment 1b ( $p < .04$ ), whereas these differences were increased in the mismatching conditions, but without reaching significance ( $p > .30$  and  $p = .07$  for nonidentical and identical mismatching, respectively). Therefore, the pattern of data reported for the solving task in the present experiment suggests that facilitation effects might be more likely than interference effects. Nevertheless, it is necessary to consider that the overall response times were higher in Experiment 2 (2,899 ms) than in Experiment 1b (2,169 ms), suggesting that the figure discrimination task itself somehow interfered with the problem-solving task. Thus, a new control experiment was designed to eliminate this bias by introducing a discrimination task not related to magnitudes.

### Experiment 3

#### Method

##### Participants

A group of 25 psychology undergraduates participated in this experiment.

##### Materials and procedure

Experiment 3 was identical to Experiment 2, except that the sequences of figures were replaced by sequences of colors. Participants indicated whether two colors were the same or different. To eliminate any reference to magnitude, colors were presented on the full screen.

#### Results and discussion

Answers to the discrimination task were not analyzed. Mean response times and error rates in the problem-solving task were analyzed in a 2 (size: small sum vs. large sum)  $\times$  2 (consistency: CL vs. IL) repeated measures ANOVA.



## Errors

The analysis indicated a main effect of consistency [ $F(1, 24) = 10.53, MSE = .017, p < .0002, \eta_p^2 = .30$ ]; participants made more errors on IL than on CL problems (.10 vs. .02, respectively).

## Response times

A total of 3.2 % of the response times were excluded as outliers or due to voice-key errors. We found a main effect of size [ $F(1, 24) = 78.19, MSE = 364,926, p < .0001, \eta_p^2 = .76$ ]; word problems with small operands (2,506 ms) were solved more quickly than those with large operands (3,574 ms). Furthermore, the main effect of consistency was significant [ $F(1, 24) = 65.68, MSE = 140,878, p < .0001, \eta_p^2 = .73$ ], showing that participants solved CL problems (2,712 ms) more quickly than IL problems (3,326 ms). These results are similar to those found in Experiment 1b, except that the overall response times were higher in this experiment, which confirms that including a same–different task between the two sentences of the word problems somehow interferes with problem solving. When compared with the results of Experiment 2, the results of this experiment support the assumption that facilitation effects in the matching condition are more likely than interference effects in the mismatching conditions. A combined analysis of the results of Experiments 2 and 3 indicated that the differences between CL and IL problems were significantly reduced in Experiment 2 relative to Experiment 3 in the matching condition ( $p < .005$ ), whereas there were no significant differences between the data reported in the mismatching conditions in Experiments 2 and 3 ( $ps > .30$ ).

## General discussion

The present study was conducted to investigate whether, during arithmetic word-problem solving, solvers construct a magnitude-based mental representation that goes beyond a conceptual representation in the form of propositions. Participants solved word problems and performed a figure discrimination task. After reading a sentence that described the comparison between the quantities of two protagonists' objects, participants were slower to discriminate as different two figures if the relationship between them, in terms of magnitude, mismatched the relationship between the quantities implied in the sentence. These findings support the hypothesis that solvers construct a mental representation that represents the situation described in the text of the problem (Thevenot, 2010). However, the present study advances beyond the previous findings by demonstrating that solvers construct a magnitude-based mental simulation and that this

occurs during online problem comprehension, possibly via an automatic activation process. It could be argued that the activation of this mental representation occurs under some level of strategic control, since each sentence describing a comparison between quantities was followed by a discrimination task. Thus, participants could have anticipated the sequence of figures during relational sentence reading. Nevertheless, this explanation is unlikely, since the use of a same–different task ensured that the figure discrimination was irrelevant for the problem-solving task. In fact, participants reported that there was no relation between the two tasks.<sup>2</sup>

These results fit nicely with those studies that have suggested that the representation of numerical magnitude may rely on a format also shared by other, nonnumerical magnitude dimensions (Cohen Kadosh et al., 2008; Walsh, 2003). According to this argument, individuals possess one, unified magnitude representation system serving diverse quantifiable dimensions. Following this notion of a shared magnitude code, the results from the discrimination task in the present study could be interpreted in terms of “within-magnitude priming” (Walsh, 2003, p. 487)—that is, a cross-domain compatibility effect between the numerical task (i.e., the numerical dimension) and the discrimination task (i.e., a nonnumerical dimension) due to the magnitude code shared by the two tasks. Thus, since participants responded more slowly in the mismatching than in the matching condition, the processing of the relational sentence of the problem activated a magnitude-based representation that interfered with the discrimination task.

It could be argued that participants did not construct a real magnitude representation, since the figures in the discrimination task only depicted the relationship between the variables contained in the relational sentence (e.g., John > Peter), but not their relative magnitudes (i.e., the ratio of the sizes of the rectangles as a function of the numbers presented in the word problem).<sup>3</sup> This is a reasonable concern, because the numerical distance effect (i.e., participants take more time when comparing digits representing closer rather than more distant quantities) indicates that numerical magnitude is processed in a “refined” way (Tzelgov, Meyer, & Henik, 1992), which would support a relative magnitude representation. However, the numerical magnitude may also be processed in a “crude” manner, such as small–large (e.g., in the numerical Stroop task; see Cohen Kadosh, 2008; Tzelgov et al., 1992). In this context, we assume that crude processing is dominant during the construction phase of the mental representation of the problem, because the next stage involves developing a plan for solving the problem (i.e.,

<sup>2</sup> Nevertheless, following a procedure proposed by Stanfield and Zwaan (2001), a post-hoc analysis showed similar match/mismatch effects in the first and second halves of the experiment.

<sup>3</sup> The authors thank an anonymous reviewer who suggested this issue.

selecting a solution procedure; see Mayer, 1985), which requires only a representation depicting the relationship between the elements of the situation described in the problem. Whether more refined processing of numerical magnitude emerges later is an open question that will need further research.

Given that we assume that the mental representation preserves only the relationship between the elements of the situation described in the text, but not its exact wording, the relationships John > Peter and Peter < John would both account for the mental representation created by the problem in Fig. 1. Hence, it could be argued that the nonidentical mismatching condition (SMALL–LARGE) would also match the mental representation activated from the relational sentence, since the first rectangle (SMALL) might be mapped onto the second person named in the word problem. Considering the example in Fig. 1, both matching (LARGE–SMALL: John > Peter) and mismatching (SMALL–LARGE: Peter < John) conditions would then reflect the situation described in the problem. In fact, Thevenot (2010) provided evidence for this possibility. Her results suggest that our nonidentical matching and nonidentical mismatching conditions might depict the situation described in the relational sentence, since only the relationship between the elements is preserved. Nevertheless, whereas Thevenot's study focused on relatively long-term representations, with participants performing recognition tasks, the present study has focused on mental representations activated during online processing of the problem. In this sense, our results support an (immediate) representation that preserves the order of variables as they are presented in the word problem (i.e., the first rectangle is mapped onto the first person named in the word problem, and the second rectangle is mapped onto the second person named in the problem).<sup>4</sup> This is consistent with the proposal of Jahn, Knauff, and Johnson-Laird (2007) that individuals construct mental models from an assertion (e.g., “The lamp is next to the magazine”) working from left to right, so that the objects are ordered from left to right in the order in which they are referred to in the assertion (i.e., LAMP–MAGAZINE instead of MAGAZINE–LAMP).

Our study also adds to previous findings by demonstrating that supporting the construction of a mental representation during processing of the word problem helps solvers carry out the solution process. Participants were faster and less error-prone in the solving task when figures matched than when they mismatched the mental representation constructed from the relational sentence. In

addition, these matching–mismatching effects were larger for IL than for CL problems. This result would be explained by the fact that, since solvers are guided by the goal of solving an arithmetic problem, they need to keep a representation of the relationship between the two variables included in the relational sentence active in working memory. Thus, when the remaining information of the problem is read, the solver must tie the known and unknown quantities to the variables represented and plan his or her solution on such a representation. In this way, a mental representation of magnitude would aid in this process, allowing the solver to quickly tie the information being processed to the information held in working memory. Since IL problems require more processing for solvers (see above), an external representation depicting the relation stated in the relational sentence, in terms of magnitude, would reduce the demands on working memory during problem solving. This finding suggests useful educational implications, since it gives us clues about how to improve the format in which word problems are presented, especially considering that word problems are often accompanied by illustrations that have little or nothing to do with problem solving. Presenting the problems accompanied by external representations reflecting the situation described in terms of magnitudes might help children to integrate the information into a coherent mental representation (Múñez, Orrantia, & Rosales, *in press*; Orrantia, Tarín, & Vicente, 2011; Vicente, Orrantia, & Verschaffel, 2008).

In sum, the results of the present study with both a discrimination and a problem-solving task provide converging evidence for the assumption that, during problem solving, solvers construct a mental representation whose structure is analogous to the relational structure of the situation described in the problem. In addition, we assume that its nature is based on magnitudes; that is, solvers mentally represent (or simulate) in terms of magnitudes the relationship between the quantities described in the situation that is represented. These findings add to a growing body of literature that emphasize the experiential view of language comprehension, according to which comprehension entails establishing an analog relationship between the text and the reader's perceptual experiences. There is converging evidence that perceptual information such as shape, orientation, or color is activated while people process words, sentences, or texts (see Zwaan, 2004). This activation reflects a mental simulation associated with experiential traces of the referential situation that an individual has encountered in the past. The present study shows that perceptual information, in this case based on magnitudes, is routinely activated when word problems are processed, since magnitude is a characteristic dimension of information that a word problem refers to.

<sup>4</sup> In fact, a closer look at Thevenot's (2010) data reveals that the direction of the relationship mentioned in the wording (i.e., our matching condition) was more often recognized as the resulting mental model created from the problem.

**Author note** This research was supported by Grant PSI2011-27737 to J.O. from the Ministerio de Ciencia e Innovación in Spain and by Grant from Junta de Castilla y León and European Social Fund to D.M.

## References

- Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews Neuroscience*, *9*, 278–291.
- Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44*, 75–106.
- Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences*, *22*, 577–660. doi:10.1017/S0140525X99002149
- Brannon, E. M. (2006). The representation of numerical magnitude. *Current Opinion in Neurobiology*, *16*, 222–229.
- Campbell, J. I. D., & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology. General*, *130*, 299–315.
- Cohen Kadosh, R. (2008). The laterality effect: Myth or truth? *Consciousness and Cognition*, *17*, 350–354.
- Cohen Kadosh, R., Lammertyn, J., & Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Progress in Neurobiology*, *84*, 132–147.
- Gebuis, T., Cohen Kadosh, R., de Haan, E., & Henik, A. (2009). Automatic quantity processing in 5-year olds and adults. *Cognitive Processing*, *10*, 133–142. doi:10.1007/s10339-008-0219-x
- Glenberg, A. M., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Processes*, *28*, 1–26.
- Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology*, *87*, 18–32.
- Jahn, G., Knauff, M., & Johnson-Laird, P. N. (2007). Preferred mental models in reasoning about spatial relations. *Memory & Cognition*, *35*, 2075–2087.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: The construction-integration model. *Psychological Review*, *95*, 163–182. doi:10.1037/0033-295X.95.2.163
- Kintsch, W., & Greeno, J. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, *92*, 109–129.
- Kintsch, W., & van Dijk, T. A. (1978). Towards a model of text comprehension and production. *Psychological Review*, *8*, 363–394.
- Lee, K., Lim, Z. Y., Yeong, S. H. M., Ng, S. F., Venkatraman, V., & Chee, M. W. L. (2007). Strategic differences in algebraic problem solving: Neuroanatomical correlates. *Brain Research*, *1155*, 163–171.
- Lewis, A. B., & Mayer, R. E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational Psychology*, *79*, 363–371.
- Mayer, R. E. (1985). Mathematical ability. In R. J. Sternberg (Ed.), *Human abilities: An information processing approach* (pp. 127–150). San Francisco: Freeman.
- Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*, 1519–1520. doi:10.1038/2151519a0
- Múñez, D., Orrantia, J., & Rosales, J. (in press). The effect of external representations on compare word problems: Supporting mental model construction. *Journal of Experimental Education*.
- Orrantia, J., Rodríguez, L., Múñez, D., & Vicente, S. (2012). Inverse reference in subtraction performance: An analysis from arithmetic word problems. *Quarterly Journal of Experimental Psychology*, *65*, 725–738.
- Orrantia, J., Rodríguez, L., & Vicente, S. (2010). Automatic activation of addition facts in arithmetic word problems. *Quarterly Journal of Experimental Psychology*, *63*, 310–319.
- Orrantia, J., Tarín, J., & Vicente, S. (2011). The use of situational information for word problem solving. *Infancia y Aprendizaje*, *34*, 81–94.
- Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences*, *14*, 542–551.
- Riley, N. S., Greeno, J., & Heller, J. I. (1983). Development of children's problem solving ability in arithmetic. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–196). New York: Academic Press.
- Schwarz, W., & Ischebeck, A. (2003). On the relative speed account of the number-size interference in comparative judgment of numerals. *Journal of Experimental Psychology. Human Perception and Performance*, *29*, 507–522.
- Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*, 153–156.
- Thevenot, C. (2010). Arithmetic word problem solving: Evidence for the construction of a mental model. *Acta Psychologica*, *133*, 90–95. doi:10.1016/j.actpsy.2009.10.004
- Tzelgov, J., & Ganor-Stern, D. (2005). Automaticity in processing ordinal information. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 55–67). New York: Psychology Press.
- Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 166–179. doi:10.1037/0278-7393.18.1.166
- Verschaffel, L. (1994). Using retelling data to study elementary school children's representations and solutions of compare problems. *Journal for Research in Mathematics Education*, *25*, 141–165.
- Verschaffel, L., De Corte, E., & Pauwels, A. (1992). Solving compare problems: An eye movement test of Lewis and Mayer's consistency hypothesis. *Journal of Educational Psychology*, *84*, 85–94.
- Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Amsterdam: Swets & Zeitlinger.
- Vicente, S., Orrantia, J., & Verschaffel, L. (2008). Influence of mathematical and situational knowledge on arithmetic word problem solving: Textual and graphical aids. *Infancia y Aprendizaje*, *31*, 463–483.
- Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*, 483–488. doi:10.1016/j.tics.2003.09.002
- Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*, 1–11.
- Zwaan, R. A. (1999). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes*, *28*, 81–88.
- Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 35–62). San Diego: Academic Press.

Copyright of Memory & Cognition is the property of Springer Science & Business Media B.V. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.