



Pictorial representations of simple arithmetic problems are not always helpful: a cognitive load perspective

Ernest C. D. M. van Lieshout¹  · Iro Xenidou-Dervou²

© The Author(s) 2018. This article is an open access publication

Abstract At the start of mathematics education children are often presented with addition and subtraction problems in the form of pictures. They are asked to solve the problems by filling in corresponding number sentences. One type of problem concerns the representation of an increase or a decrease in a depicted amount. A decrease is, however, more difficult to represent in paper pictorially than an increase. Within the Cognitive Load Theory framework, we expected that commonly used decrease problems would be harder and slower to solve than increase problems because of higher intrinsic cognitive load. We also hypothesised that combining the pictorial information with auditory information would reduce this load and, as a result, would lead to improved performance. We further expected that these effects would be most prominent in children who score below average on a general mathematics test. We conducted an experiment with sixty children attending the first grade of primary school, who were divided into a higher and lower mathematics-achieving group. As expected, children performed worse on the decrease problems compared to the increase problems. Also, the combination of pictorial and auditory information reduced the accuracy lag of the decrease problems compared to the increase problems. With response time this effect occurred only in the group of lower mathematics achieving children. These results are in line with the cognitive load theory framework, but we also offer an alternative explanation regarding attention.

Keywords Arithmetic · Pictures · Cognitive load theory · Modality effect · Mathematical word problems · Primary school

✉ Ernest C. D. M. van Lieshout
E.C.D.M.van.Lieshout@vu.nl

¹ Department of Educational and Family Studies, Section Educational Neuroscience, Faculty of Behavioural and Movement Sciences and LEARN! Institute, Vrije Universiteit Amsterdam, van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands

² Mathematics Education Centre, Loughborough University, Loughborough LE11 3TU, UK

1 Introduction

At the start of mathematics education children are taught first to understand the meaning of addition and subtraction; later, they start learning how to formally represent these operations in the form of number sentences ($a \pm b = c$). Many mathematics curricula, in the Netherlands and outside, contain exercises on writing number sentences. These are available in worksheets, practice and testing booklets, on paper, or on the computer, often in the form of pictures. Practicing how to write number sentences can have several functions: First, it elucidates the formal representational function of the number sentence for a mathematical problem that is presented in a verbal or pictorial context. Secondly, it helps the student get acquainted with the use of symbols such as the plus, minus and equals signs and the place of all elements, including the numbers. Thirdly, it is a form of exercising basic arithmetic operations and it has been shown that practicing to write number sentences for arithmetic word problems can help a child learn how to solve them (Carpenter, Moser, & Bebout, 1988; Stellingwerf & van Lieshout, 1999).

Mathematical word problems can be used to represent real life situations helping the child: 1) Understand the connection between the represented situation and the mathematical operation, and 2) exercise the application of mathematical operations in real-life situations. However, many of the word problems used in school are not really inviting the student to apply real-world knowledge due to their artificial nature (e.g., Verschaffel, De Corte, & Lasure, 1994) and are, therefore, not always the best vehicle to represent real-life situations. Still, current maths curricula aim at translating real-life situations into mathematical problems to make maths more realistic. Naturally, this is a difficult process. In a concrete, real-life quantitative problem, a problem solver is perhaps able to see the elements of the involved sets of objects or measure the quantities and manipulate them in order to carry out operations like adding and subtracting. In contrast, in the case of a mathematical word problem, which can be seen as the written or oral description of the real-life problem, the solver has to build an internal representation – a mental model – based on the textual description (Kintsch, 1986; Thevenot, 2010). Therefore, solving a simple word problem needs extra skills that are not required for a concrete real-life problem situation (Rasmussen & Bisanz, 2005). In other words, there are now extra steps requiring – amongst other cognitive abilities – text comprehension and – in the case of written text – word recognition abilities (Fuchs, Fuchs, Compton, Hamlett, & Wang, 2015). It is believed that by using pictures instead, it helps overcome this reliance on linguistic skills by making the problem more concrete or realistic. Inspecting mathematics curricula, one notices the frequent use of such pictorial problems. However, Dewolf, Van Dooren, and Verschaffel (2016) did not find a positive effect of pictorial illustrations on the performance of 9–12 years old students. Actually, Berends and van Lieshout (2009) showed that adding a picture to an arithmetic problem could even be detrimental to the performance of about 9.5 years old students. In the present study we focussed on even simpler problems, which are used in grade 1. These problems consist of pictures that depict the problem situation and there is no accompanying text, since children are not proficient enough to read mathematical problem texts at this age. To our knowledge, no systematic research has been carried out on pictorial representations of simple addition and subtraction problems. The present study helps build a theory of the effects of pictorial problems and the way they can be best used in exercises and teaching.

The types of pictures that are commonly used to depict addition or subtraction are diverse. They appear to resemble the semantic structures of the types of problems that are distinguished

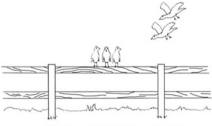
in the mathematical word problem literature (Carpenter & Moser, 1984; De Corte, Verschaffel, & De Win, 1985). One of the formats is the “change” type, which has a start set, a change set, which is either an increase or decrease of the start set, and an end set. In the “combine” type, either the whole or one of the parts is the unknown. Another example is the “compare” type in which one set contains more or less objects than the other set. Sometimes also a fourth type is distinguished: the “equalise” type, which is a combination of the change and compare type. Carpenter and Moser (1984) and De Corte et al. (1985) have shown that the semantic structure of an arithmetic word problem strongly determines the solution strategy and the level of difficulty of the problem. In addition, the position of the unknown in a word problem text also contributes to its semantic structure.

In the present study, we focussed on the pictorial version of the dynamic change problem type where an increase or decrease of a certain amount is visualised and where the unknown is either the last set – in the case of the increase problems – or the first set – in the case of the decrease problems. The terms ‘increase’ and ‘decrease’ are used here to describe what kind of situation the picture tries to depict and not necessarily the operation that has to be carried out to solve the problem. This dynamic situation must be represented by the developers of the material in a picture. This picture is, of course, at least if it is on paper, static, however, the problem situation is dynamic. Thus, it is important that the picture demonstrates the dynamic change in terms of depicting the start or end set, the amount of change, and its direction (increase or decrease). Figure 1 gives examples of such problems, which we will call henceforth *increase* and *decrease problems*.


As shown in Fig. 1, the increase situation problems show the two subsets, i.e., the start set (the augend) and the change set, i.e., the increase (the addend), which have to be added up to find the unknown total set (the sum). Expressed into a number sentence, this would be stated as “ $a + b = ?$ ”. In contrast, the decrease situation problems show one of the subsets, namely the decrease (subtrahend), and the end set (the difference), which have to be *added* to find the unknown start set (the minuend). In this case, the number sentence that describes the situation is “ $? - b = a$ ”. So, although in the decrease situation the picture suggests a decrease and the empty number sentence shows a minus sign, the required operation is actually addition ($a + b = ?$), not subtraction.

In the increase situation (Fig. 1) the two subsets are not yet joined but seem to be going to be joined. This agrees with the child’s task. However, for the decrease situation, the pictures show the result of the already carried out separation of the minuend into a subtrahend and a difference, which makes the yet to be carried out correct action less obvious. Finding this unknown means that the child should inhibit the possible tendency to consider the minus sign as an instruction to subtract the two given sets from each other – which could be carried out but would lead to an incorrect result. Instead the child must understand that the unknown start set has to be reconstructed by an addition because this unknown set represents the whole of two parts (the subtrahend and the difference). In other words, the child has to understand that the static picture represents the final stage of a dynamic change of the minuend into a subtrahend and the resulting difference, which occurred in the past. Thus, it can be assumed that the pictorial decrease problems probably pose a larger cognitive load than the increase problems. This would not be due to the fact that subtractions are more difficult than additions (Campbell, Fuchs-Lacelle, & Phenix, 2006), because subtraction is not actually needed in this case. In terms of the Cognitive Load Theory (CLT, Sweller, 1994), the decrease problems are characterised by more element interactivity because processing only the two set sizes as information elements does not take into account the element of the outward movement of

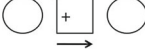
Examples of increase problems
(Solution: 5, Number sentence: $3 + 2 = 5$)

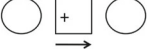


Three birds are sitting on a fence. Then two birds fly in. How many birds are sitting on the fence in total?




There are three children standing on the ship. Two children are walking to them. How many children are then on the ship in total?







Examples of decrease problems
(Solution: 5, Number sentence: $5 - 2 = 3$)



Two birds fly away. There are three birds still sitting on the fence. How many birds were sitting on the fence in the beginning?



Two children walk away from the ship. There are three children still standing on the ship. How many children were standing on the ship in the beginning?






Fig. 1 Examples of the experimental problems: Two pictorial increase and decrease problems, their auditory-presented verbal analogues (original in Dutch), the correct solution and the required schematic number sentence (i.e., circles and square). The empty number sentence was presented in every condition. The examples are based on real problems used in practice and testing booklets in school. Note that the texts were never visible to the participants; they were spoken out loud by the experimenter. All experimental pictures were drawn by the first author

objects (e.g., birds or children), which indicate that the situation was different before. Also the minus sign in the decrease problems could add to this element interactivity because it does not point to the required correct operation (i.e., addition). According to the CLT, higher element interactivity leads to higher intrinsic load. The present study examined why pictorial information can sometimes be unhelpful in mathematics education and how its efficacy could be improved. Specifically, our research question was whether CLT could explain the possible difference between performance on the increase and the decrease problems. As described earlier, we had hypothesised that decrease problems would be more difficult than increase problems.

Based on CLT (Sweller, 1994), cognitive load can be reduced by presenting the information not just in one modality, i.e., solely in the visual or auditory modality, but simultaneously in both (Sweller, van Merriënboer, & Paas, 1998). This way, the available working memory capacity that can be used for the task in hand is raised. Also, according to the ‘modality effect’¹ in the cognitive theory of multimedia learning, a student learns better from a text with pictures when the text is presented auditory instead of visually (Moreno & Mayer, 1999). The idea of decreasing cognitive load by distributing it across different modalities is based on Baddeley’s (2009) Working Memory (WM) model, which assumes two subordinate systems, the visuo-

¹ In the present study, we use this term to examine the effect of the combination of pictorial and auditory information.

spatial sketchpad for the temporary storage of visual and spatial information, and the phonological loop for the temporary storage of acoustic and phonological information. The model further contains a supervisory system, the central executive, which processes the incoming and stored information. So, Baddeley's model contains one limited-capacity component for auditory and one for visual information. Based on this model, CLT assumes that more cognitive load can be processed when both systems are used simultaneously during task implementation.

Therefore, we expected that children's pictorial problem-solving performance would improve when accompanied by an auditory description of the problem. More importantly, we expected to find this effect more in the case of the decrease problems compared to the increase problems due to the extra cognitive load they impose. To rule out the alternative explanation that perhaps the auditory information lead to improved performance on its own, the spoken texts were also presented as a stand-alone control condition. So, the experimental design consisted of three modality conditions: pictorial, auditory and their combination. Finding a reduction in the difference in performance between the increase and decrease problems in the combination condition compared to the two pure modality conditions, would support the hypothesis that decrease problems are more difficult than increase problems due to the extra cognitive load.

WM is one of the factors that explain individual differences in mathematics, i.e., poor mathematics performers have lower WM capacity (Passolunghi, Vercelloni, & Schadee, 2007; Xenidou-Dervou, Molenaar, Ansari, van der Schoot, & van Lieshout, 2017). This could make it more difficult for them to deal with the extra cognitive load that is caused by trying to understand that the static picture of a decrease represents a change that has already happened. When comparing expert and novice mathematicians, Stylianou and Silver (2004) found that novices view their own graphical representations of a maths problems more as a static object, whereas experts as more of a dynamic object. Therefore, we expected that children with lower mathematics achievement would profit more from a combined modality presentation compared to children with higher mathematics achievement.

2 Method

2.1 Participants

Sixty children (29 boys and 31 girls mean age, mean age: 6.70, $SD = 0.55$) from a Dutch urban primary school participated in the study after having given informed consent. We wanted to make a distinction between high and low mathematics achievers. Due to the school's privacy policy we had no access to the children's individual mathematics ability scores, which could form the basis for determining the child's level. Instead, a school staff member selected 20 children from each of the school's three grade 1 groups. This person was requested to choose in each class, four subgroups of five children, who represented four levels of mathematics achievement as distinguished by the standardised scores on the national general maths ability test in the Netherlands (Cito; Janssen & Kraemer, 2002). This resulted in 15 children in each mathematics achievement level. The four levels represent the four quartiles in achievement levels in the Dutch population. The two highest and the two lowest levels were combined in the analyses resulting in a low and high maths-achievement group (30 children each). Table 1 contains the information regarding the participants. A χ^2 -test showed that the distribution of gender across the two mathematics levels did not differ, $\chi^2(1, N = 60) = 0.07, p = .796$. The

mean age across the groups was also similar, $t(58) = 0.32$, $p = .751$. To further establish the high- and low-maths achievement division, we also assessed the children on an addition and a subtraction speed test. Both differences in mean score between the two groups (Table 1) were significant, respectively, addition: $t(58) = 2.79$, $p = .007$ and subtraction: $t(38.49) = 3.30$, $p = .002$, corroborating the division into the two general maths achievement levels.

2.2 Design and procedure

The study had a factorial design with two within- and one between-subjects factors. A within-subjects factor is a factor that is varied in the same way within each participant. A between-subjects factor is varied across different groups of participants. The within-subjects factors concerned the change (increase or decrease) in the size of a depicted or mentioned amount in the problem, which we will call henceforth 'size change', and the presentation modality (pictorial, auditory or the combination of both). The between-subjects factor concerned the two maths achievement groups (low and high). The children participated in three individual testing sessions within a month (close to the end of the school year). Each session contained the problems of one presentation modality and, at the start of the second session also the two arithmetic speed tests.

2.3 Material

2.3.1 Experimental problems

During each trial a laptop screen showed in PowerPoint a diagram with an empty number sentence and a plus or minus sign. It also showed, in the pictorial and combination condition – not in the auditory condition – an increase or decrease picture (see Fig. 1). Between each trial, a fixation cross appeared in the middle of the screen. Children were instructed to look at the cross. When they did, the experimenter started the presentation of the next trial. The child sat next to the experimenter so that both could easily see the screen. Children had to point to the squares and circles in the order of the number sentence and tell the experimenter which number belonged in which circle and square. In the auditory and combination conditions, the child was asked to start with these actions only after the experimenter had finished reading the text, which the child could not see. During the instruction phase, the experimenter presented the pictorial or the auditory practice problems. She would say that birds are going to sit down on the fence or fly away and that children are walking up to or down from a ship. When the child did not understand the instruction, the experimenter pointed out to the event that took place in terms of the amount of change and the parts of the total set. While the child solved the

Table 1 Numbers of children, age and speed test performance in the high and low maths-achievement group

Maths level	<i>n</i>		Age		Speed test			
					Addition		Subtraction	
	Boys	Girls	<i>M</i>	(<i>SD</i>)	<i>M</i>	(<i>SD</i>)	<i>M</i>	(<i>SD</i>)
High	15	15	6.68	(0.63)	6.47	(2.19)	5.30	(2.51)
Low	14	16	6.72	(0.46)	5.10	(1.54)	3.67	(1.03)

problems, the experimenter used a scoring form where she indicated whether the child mentioned the size of the unknown total set correctly and whether the complete number sentence was correct.

In the pictorial condition, RT measurement (with a stopwatch) started the moment the picture was presented. In the auditory condition, however, RT measurement started right after the experimenter finished the auditory description (≈ 10 s). That is because, in the pictorial condition all the necessary information was available there and then in the picture itself, however, in the auditory condition one had to first hear the entire description before having all the necessary information to start solving the problem. RT measurement in the combined condition was the same as in the auditory one, i.e., it started right after the experimenters' description ended. This way, RT measurement started in all conditions at the time all the to be presented information was available to the participants. In all conditions RT measurement stopped after the child completed the number sentence. Thus, we could not directly compare the pictorial with the other conditions in terms of RT. However, that was not a problem for our research question, which mainly concerned comparing the increase with the decrease problems. Both problem formats (increase and decrease) contained all modality conditions (pictorial, auditory and combination). When calculating RT differences between the problem formats, the extra time in the auditory and combination conditions cancelled each other out. This was also true for the interactions of size change with the other factors (modality and maths achievement group). In all conditions RT measurement was stopped after the child had completed the number sentence.

Each participant was presented with 36 problems, divided into 12 per modality condition. Half of these 12 problems concerned an increase and the other half a decrease. Half of the increase problems and half of the decrease problems represented a ship with children and the other half, a fence with birds (Fig. 1). Ships and fences were alternated across the problems. To counterbalance order effects in the size change factor, we used an ABBA scheme within each modality condition, where A indicates an increase and B indicates a decrease problem. In order to minimize the possibility that the participant could predict whether the next problem would concern an increase or decrease, we in fact used two series of an ABBA order and one series of a BAAB order. The order of these three series was rotated across the three sessions. This resulted in the orders ABBA – BAAB – ABBA for the first, ABBA – ABBA – BAAB for the second and BAAB – ABBA – ABBA for the third presented modality condition. For the six problems in both size change situations (increase or decrease), six combinations of numbers were used: (3, 1), (3, 2), (4, 3), (5, 1), (5, 2) and (6, 2). These were the numerosities that were visible as sets in the pictures or spoken in the auditory condition. For example, the combination (3, 1) in an increase situation meant that child should consider it as the addition $3 + 1 = 4$. But in a decrease situation, it meant $4 - 1 = 3$. In both cases 4 (the sum in the former and the minuend in the latter case) was the desired solution for the unknown. Each number pair was used an equal number of times in the increase and decrease problems. The only difference between an increase and a decrease problem with the same number pair and the same picture (birds/children) was the direction of the movement of the to be added or separated amount. The sum of the number combinations was never higher than eight. In the decrease problems, the number of children or birds that were left was always larger than the number that went away. The reverse situation (e.g., three birds flying away while two remain) could have given the child the idea that filling in the number sentence with ' $2 - 3$ ' would be incorrect. Within the two change-situations the six number combinations were distributed pseudo randomly. We prevented two consecutive trials containing the same combination of numbers.

The depicted direction of the movement of the children or birds in the pictorial problems signified whether the problem concerned an increase or a decrease. In the auditory problems, the direction was expressed in words. The operation sign (+ or -) in the schematic empty number sentence could also act as a cue for the direction of the change. We used exactly the same schematic empty number sentence and the same increase or decrease structure as in the problems encountered in the Dutch mathematics curriculum Pluspunt (Pap, 2002); only the pictures that we used were different, which does not affect the ecological validity of our experiment. In each modality condition, the experimental problems were preceded by two practice problems. These were identical to the experimental problems, with the difference that the number combination was (2, 1). Feedback was provided only during the practice problems.

Possible order effects in the modality factor were counterbalanced by use of a Latin square design. This means that the order of the three modality conditions was systematically varied in a specific way in order to keep possible order effects the same for each condition. In one order, the first session contained the pictorial (P) condition, the second session contained the combination (C) condition and the last one contained the auditory (A) condition. In another order, the sequence was: C – A – P. Finally, there was a third order with the sequence: A – P – C. An equal number of participants was allocated to each of these three orders.

2.3.2 Other maths tests

Two arithmetic speed tests were used to further distinguish between the two maths level groups of children, one for addition and one for subtraction. The addition speed test consisted of four blocks of six addition problems. The numbers in the problems were based on the same six number combinations as the ones used in the experimental problems. The problems were pseudo randomly ordered within each block, with the restriction that the first number pair was exchanged with another pair of the remaining five pairs when one of its numbers was equal to one of the numbers of the last pair of the previous block. The subtraction speed test was developed in the same way.

2.4 Analysis

Two accuracy variables were used from the responses to our experimental problems: The proportion of correctly determined total set sizes (the unknowns) and the proportion of correctly filled in number sentences. The score on the latter reflects the stricter requirement to have all the three numbers in the number sentence filled in correctly. The first accuracy variable was the main outcome variable, but we also checked whether results corresponded for the latter. The third dependent variable was RT.

Proportions were based on the child's dichotomous response: correct or incorrect. Comparing mean proportions that differ in their distance to the extremes (0 or 1) of the distribution can violate the ANOVA assumption of homogeneous variances, because variances close to the extremes of the distribution approach zero. This is the case when the smallest or largest of two proportions in a comparison are respectively approximately lower than .3 or higher than .7 (Agresti, 2002, p. 120). This indeed occurred in our study. Therefore, Generalised Estimating Equations (GEE) were used to analyse the data (see Jaeger, 2008) with a probit link function (Garson, 2013). The probit function allows the transformed score to vary between minus infinity and plus infinity (with zero as midpoint) instead of 0 and 1.

In contrast, RT data were analysed with a mixed model ANOVA. For each participant, the median RT across the six problems was calculated.

3 Results

3.1 Accuracy

Accuracy of the calculation of the total set size (the unknown) was analysed with a 2 (Maths Level: low versus high) \times 3 (Modality: pictorial, auditory and combination) \times 2 (Size change: increase versus decrease) GEE with repeated measures on the latter two factors. Figure 2 shows the mean proportions for all conditions. The expected interaction of Maths Level \times Modality \times Size Change did not reach significance, Wald $\chi^2(2, N = 60) = 0.96, p = .619$. In contrast, the expected effect of Modality \times Size Change was significant, Wald $\chi^2(2, N = 60) = 6.89, p = .032$. To examine the simple effects within this two-way interaction, we tested the effect of size change within each of the three modalities. Within the pictorial and auditory modality the size change effect was significant, Wald $\chi^2(1, N = 60) = 12.96, p < .001$ and Wald $\chi^2(1, N = 60) = 16.06, p < .001$, respectively. There was no size change effect in the combined modality condition, Wald $\chi^2(1, N = 60) = 0.02, p = .886$. So, the effect of size change was larger in the unimodal conditions compared to the multimodal condition. Thus our hypothesis was confirmed, the accuracy lag between the decrease and increase problems was reduced when combining pictorial and auditory information.

We also tested the simple effects of modality within the size-increase and size-decrease level of the factor size change. Within the size-increase condition, we found a significant effect of modality, Wald $\chi^2(2, N = 60) = 47.48, p < .001$. Pairwise comparisons with Bonferroni correction showed that the number of correctly solved problems in the combination and pictorial condition did not differ, $p = .298$. However, performance in the combination condition was significantly better compared to the auditory condition, $p < .001$. Performance in the pictorial condition was higher than the auditory condition, $p < .001$. In the size-decrease condition, we saw a significant effect of modality, Wald $\chi^2(2, N = 60) = 113.31, p < .001$. Contrary to the findings in the size-increase condition, pairwise comparisons with Bonferroni

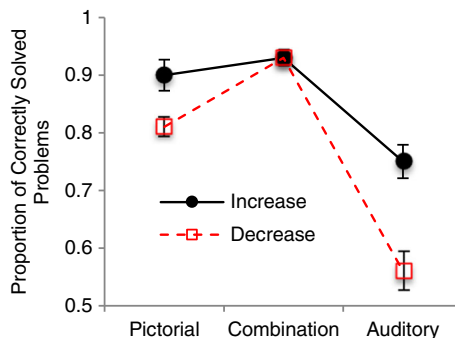


Fig. 2 Proportion of correctly solved problems (i.e., unknown total set sizes) in the three modality conditions (pictorial, auditory and combination) and the two set-size change conditions (increase versus decrease). Vertical bars depict *SEs*. The mean probit scores within the set-increase conditions were, from pictorial to auditory: 1.30 ($SE = 0.10$), 1.51 ($SE = 0.11$) and 0.66 ($SE = 0.09$). For the decrease condition the means were: 0.86 ($SE = 0.10$), 1.48 ($SE = 0.11$) and 0.16 ($SE = 0.09$). Note that theoretical probit scores vary from $-\infty$ to $+\infty$

correction showed that the difference between conditions within all the three pairs of modality factor were significant, $p < .001$. The order of conditions in terms of highest to lowest performance was as follows: combination, pictorial, auditory. The superiority of the combination condition was in agreement with our expectation that the combination of pictorial and auditory information would improve the performance.

Apart from the two-way interaction between Modality and Size Change, the main effect of maths achievement level was also significant, Wald $\chi^2(1, N = 60) = 8.49, p = .004$, in favour of the high achievers, $M_{\text{High achievers}} = .87 (SE = .02)$ and $M_{\text{Low achievers}} = .80 (SE = .02)$. The main effects of modality and size change were significant as well: respectively, Wald $\chi^2(2, N = 60) = 147.38, p < .001$ and Wald $\chi^2(1, N = 60) = 16.22, p < .001$. Because these two factors were involved in the reported Modality x Size Change interaction, they were not further interpreted.

The preceding analysis regarded data on accuracy of the solution of the unknown total set size. We also ran the analyses with the more stringent scoring of the accuracy of the filled in number sentence as dependent variable. Results showed that the Modality x Size Change interaction was not significant, Wald $\chi^2(2, N = 60) = 3.48, p = .175$, which means that the effect of size change was the same in all modalities. There was, however, a significant main effect of modality, Wald $\chi^2(2, N = 60) = 146.95, p < .001$, indicating that the three means were different across the modalities. Pairwise comparisons with Bonferroni correction showed that the difference between conditions within all the three pairs of the modality factor were significant, $p < .001$. The order of conditions in terms of highest to lowest performance was as follows: combination, $M = .91 (SE = .01)$, pictorial, $M = .84 (SE = .02)$, and auditory $M = .59 (SE = .02)$. The superiority of the combination condition was in agreement with our expectation that the combination of pictorial and auditory information would improve performance. The main effect of size change was significant as well, Wald $\chi^2(1, N = 60) = 26.40, p < .001$ in favour of the increase problems: $M_{\text{Increase}} = .85 (SE = .01)$ and $M_{\text{Decrease}} = .75 (SE = .02)$. Lastly, the main effect of maths achievement level was significant, Wald $\chi^2(1, N = 60) = 8.49, p = .004$, in favour of the high achievers, $M_{\text{High achievers}} = .84 (SE = .02)$ and $M_{\text{Low achievers}} = .77 (SE = .03)$.

In sum, with respect to the success in calculating the unknown, i.e., the sum of the two presented set sizes in both the increase and decrease problems – which was our main accuracy variable – we found that the combination of pictorial and auditory information was better than the pictorial only (and also the auditory only) condition in the decrease problems, but this did not depend on mathematical achievement level. As hypothesised, the combination condition suppressed the negative effect of the decrease problems. With the more stringent scoring of the accuracy of the filled in number sentence as dependent variable, we did not find that the increase and decrease problems differed with regard to the influence of modality. Instead, the combination of pictorial and auditory information led to the highest performance in both problem types.

3.2 RT

As pointed out earlier, the pictorial presentation mode in the analysis of RT could not be compared with the two other presentation modes, due to the difference in the RT scoring procedure. However, the comparison between the combination and the auditory condition was possible. Also, the interactions of the pictorial conditions with the other factors could be analysed. When RTs of only correct solutions were used, the RT data of only 4 out of the 60

participants were complete. To see whether using all RT data, i.e., from both correctly and incorrectly solved problems was permissible, we calculated the correlation between mean proportion accuracy and mean median RT within each of the within-subjects experimental conditions (Modality by Size Change). All six (Spearman ρ) correlations were negative with three of them significantly negative, $\rho_{\text{Size increase in Pictures}} = -.16, p = .238, \rho_{\text{Size increase in Combination}} = -.07, p = .613, \rho_{\text{Size increase in Auditory}} = -.47, p < .001, \rho_{\text{Size decrease in Pictures}} = -.23, p = .08, \rho_{\text{Size decrease in Combination}} = -.30, p = .020, \rho_{\text{Size decrease in Auditory}} = -.371, p = .004$. Therefore, there was no indication of a speed-accuracy trade-off and it was thus justified to include all RTs.

Median RTs were submitted to a 2 (Maths Level: low versus high) \times 2 (Modality: pictorial, auditory and combination) \times 2 (Size Change: increase versus decrease) ANOVA with repeated measures. Mauchly's test showed that the sphericity assumption was violated for Modality \times Size Change, $\chi^2(2) = 28.50, p < .001$. Therefore, we used the Greenhouse-Geisser correction of the degrees of freedom for the F -test. As expected, the Modality \times Size Change \times Maths Level interaction was significant, $F(1.44, 83.25) = 3.95, p = .036, \eta_p^2 = .06$ (Fig. 3). This means that the Modality \times Size Change interaction, which we already found in the accuracy data, was in the case of RT modified by the maths level. Also the Modality \times Size Change interaction, $F(2, 116) = 6.25, p = .003, \eta_p^2 = .10$ and the main effects of modality, $F(2, 116) = 25.93, p < .001, \eta_p^2 = .31$, size change, $F(1, 58) = 32.69, p < .001, \eta_p^2 = .36$ and maths level, $F(1, 58) = 10.77, p = .002, \eta_p^2 = .16$ were significant (Fig. 3). The Modality \times Maths Level interaction was marginally significant, $F(2, 116) = 2.87, p = .06, \eta_p^2 = .05$. The two-way interactions and main effects were not interpreted because they were involved in the mentioned three-way interaction.

To investigate the simple interaction effects within the three-way interaction, a 2 (Modality) \times 2 (Size Change) ANOVA with repeated measures was conducted within each of the two maths level groups. In the high maths group, Mauchly's test of sphericity was significant for Modality, $\chi^2(2) = 12.62, p = .002$ and for Modality \times Size Change, $\chi^2(2) = 6.57, p = .037$. In

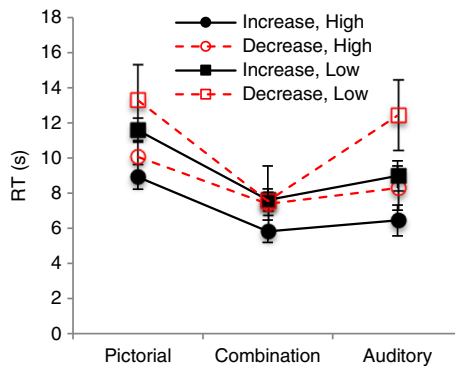


Fig. 3 Median RTs in the three modality conditions (pictorial, auditory and combination), the two set-size change conditions (increase versus decrease) and the two mathematics achievement level groups (high versus low achievement level). Vertical bars depict SEs . Note that the RT in the combination and auditory condition show the RT from the moment the experimenter was ready with reading the problem text aloud, which meant that already approximately 10s had passed since the trial started. Therefore the RTs of the three modality condition should not be compared directly (except for the auditory with the combination condition). On the contrary, RT differences between the increase and decrease problems and between maths levels can be compared across the three modality conditions and the maths achievements groups

the low maths group, it was significant for the Modality x Size Change interaction, $\chi^2(2) = 18.45, p < .001$. Consequently, the degrees of freedom of the respective F -tests were corrected.

In the high maths level group there was a main effect of size change, $F(1, 29) = 45.07, p < .001, \eta_p^2 = .61$ (the children responded faster to the increase problems than the decrease problems, $p < .001$) and modality, $F(2, 58) = 10.37, p < .001, \eta_p^2 = .26$ (although the mean RTs of the combination and auditory condition did not differ significantly, $p = .862$). The Modality x Size Change interaction was not significant, $F(1.65, 47.97) = 0.53, p = .560, \chi_p^2 = .02$. In contrast, this Modality x Size Change interaction was significant in the low maths level group, $F(2, 58) = 6.54, p = .003, \eta_p^2 = .18$. Pairwise comparisons showed that the increase problems were answered significantly faster than the decrease problems in both the pictorial, $p = .001$, and the auditory condition, $p = .005$. In the case of the combination problems, there was no significant speed difference, $p = .725$.

In sum, as hypothesised, the participants from the low maths achievement group solved increase problems faster than decrease problems in the pictorial and auditory conditions, but not in the combination condition. RT results were largely in agreement with the accuracy results, with the difference that the Modality x Size Change interaction was significant in the analysis of accuracy, whereas in the case of RT, it was only significant for the low maths group.

4 Discussion

This study examined the conditions under which pictorial problems can be detrimental or helpful for children when they solve simple increase and decrease problems. On the basis of the Cognitive Load Theory (CLT; Sweller, 1994), we hypothesised that size-decrease problems would lead to larger cognitive load than size-increase problems due to higher element interactivity. Indeed, children performed worse – both in accuracy and RT – in the decrease problems compared to the increase problems when they only had pictorial (or auditory) information at their disposal. As described in the introduction, this difference cannot be attributed to the fact that subtraction is more difficult and takes more time than addition in general (Campbell & Xue, 2001; Kamii, Lewis, & Kirkland, 2001). Actually, both problem types that we used – increase and decrease – required an addition operation to be solved correctly (Fig. 1).

Furthermore, we found a clear modality effect in these decrease problems: When the pictorial representation was accompanied by auditory information children's accuracy increased. This is in line with CLT's assumption that cognitive load can be reduced by presenting the information not just in one modality, i.e., solely in the visual or auditory modality, but simultaneously in two different modalities (Sweller et al., 1998). We want to stress that the increase in performance in the case of the pictorial condition by adding auditory support can not solely be attributed to the auditory information itself. Actually, the stand-alone auditory condition produced the worst performance. On the contrary, as was expected, it was the combination of pictorial and auditory information that led to the best performance. When the auditory and pictorial information were combined - to decrease the cognitive load imposed by these problem types - accuracy in these problems was now as high as in the increase problems.

Taken together these findings support the assumption that the decrease problems in the pictorial (and the auditory) modality imposed more cognitive load than the increase problems. In general, pictorial maths problems can impose extra cognitive load. Aside from the present

findings, this has also been shown in the case of maths problems with written text, where the cognitive load was increased by the so-called split-attention effect (Berends & van Lieshout, 2009; Leikin, Leikin, Waisman, & Shaul, 2013). The split-attention effect occurs when one must divide his or her attention between several sources of information within the same channel, like pictures and written text, which both enter via the visual channel (Sweller, 1994).

It should be noted that we analysed the accuracy data on the basis of two measures: Accuracy on the unknown and the more stringent dependent variable of accuracy on the entire filled-in schematic number sentence (see Fig. 1). In both cases results were comparable: The higher difficulty of pictorially presented decrease problems diminished when auditory support was added. The only difference was that the interaction between modality and size-change was absent in the case of the entire number sentences' accuracy. In other words, the difference in accuracy between increase and decrease problems did not depend on the specific modality in the case of the more stringent dependent variable. However, this interaction did appear in the RT analysis and the used RT measurement did concern the completion of the entire number sentence.

Analyses on the accuracy and RT data showed largely the same results, with the exception that with the RT data, the modality effect on the performance difference between the increase and decrease problems only occurred in the low maths achievement group. This group of children, as hypothesised, profited more – speed-wise – from the combination of modalities compared to the higher achievers in mathematics. As we described earlier, poor performers in mathematics, often also perform poorly on WM tests (Passolunghi et al., 2007; Xenidou-Dervou et al., 2017). Evidently, the combination of pictorial and auditory information compensated for lower WM capacity.

Our starting point was the CLT, which assumes higher intrinsic cognitive load when there is high element interactivity in a problem. However, one may consider an alternative account for our findings. The increase picture shows what is about to happen, whereas the decrease problem shows what happened (Fig. 1). It is possible that the mental representation of what the picture depicts in terms of what is happening, may be more dependent on recognising the elements in the picture that help to mentally represent these different events than on effectively storing the presented information in WM. Perhaps children use the plus and minus sign in the number sentence as a cue for the required operation. In the increase problems, the meaning of the plus sign coincides with the required addition of the two visible sets. In the case of the minus sign, however, it does not; the child must identify the necessary elements while ignoring the sign. On the other hand, in the combination condition, the auditory information may perhaps help direct the child's focus to the critical details in the picture (e.g., birds flying away). If the experimenter said, e.g., 'Two birds fly away', the yet unnoticed flying direction of the birds probably becomes obvious to the child. Based on the CLT explanation, in the pure pictorial problem, the information concerning the direction of the movement and the contradicting minus signs gives rise to extra load because it has to be processed and temporarily stored in WM. But, based on the alternative explanation, this information will not be processed and stored because the suggested movement was not noticed. However, there is in fact little support for this attention explanation. If the children had ignored this information in the pure pictorial decrease problems, one would not expect slower RTs than in the case of the pure pictorial increase problems. So, the CLT explanation seems still the best one of the two.

Future research could disentangle these two explanations by assessing children's eye fixations in these different conditions. In Dewolf, Van Dooren, Hermens, and Verschaffel (2015) eye-tracking experiment participants did not even look at the illustration that

accompanied the text. In our experiment, there was no written text in any condition. Moreover, it is unlikely that the children in the present experiment did not look at the pictures since they scored above chance level. Nevertheless, it is possible that they did not always look carefully enough at the illustration and therefore did not notice the children's or birds' relevant "movement". Thus, an eye tracking experiment could elucidate children's problem-solving process.

Of course, we do not know exactly what kind of strategy the children actually used to solve the decrease problems. One possibility is that they tried to directly find the minuend (i.e., the "?" in the number sentence " $? - a = b$ "), either before or after filling in the known numbers in the number sentence schema. They could have done this by, for example, starting to count the objects (the birds or children) in one of the two subsets (i.e., the subtrahend, "a", or the difference, "b") and by continuing the count in the other subset. Alternatively, they could have used number facts instead. Another possibility is that they used, what De Corte and Verschaffel (1981) have called, an 'intelligent trial and error strategy'. This means that children first try to guess the number that should be filled in. When they subsequently test their answer by carrying out a subtraction that does not result in a difference that is equal to the one in the end set, they would then make a new attempt. Earlier we explained how lower performance in the decrease problems could not be explained by the process of applying subtraction. Using the subtraction procedure during the trial and error strategy could indeed decrease performance because subtraction is more difficult than addition (Campbell & Xue, 2001; Kamii et al., 2001). But carrying out repeated attempts to find a suitable solution by trial and error, would also take extra time. This increase in RT would not occur simply because of possibly carrying out the subtraction; instead it would occur because of the multiple attempts that were brought about by the high cognitive load. In future research, it would be interesting to examine the mentioned solution strategies. In the introduction we suggested that, if children make a mistake on a pictorial decrease problem, they perhaps failed to inhibit the tendency to subtract - instead of adding - the two numbers. So, future research could also examine the association between the kind of errors children make and their inhibition abilities.

Examining performance in such basic problem types at the early years of schooling is important because it can have implications for their future mathematics development, especially for children who face difficulties with arithmetic. Thus, early difficulties with decrease problems could hinder further development to solve more complex problems. Jenks, de Moor, and van Lieshout (2009), for example, demonstrated in a longitudinal study that a group of children who performed poorly on a simple number fact task, could catch up with the higher performing group after some years, but at the next developmental stage fell behind on more complex calculation task. Future research could examine whether performance lags in simple problems such as the ones used in the present study, have consequences for learning to solve more complex problems.

The aim of the developers of this kind of problem is probably to devise pictures that constitute snapshots of a changing situation, i.e., an increase or decrease, in order to make the problem situation more realistic and therefore also more understandable. But, as pointed out above, in the case of a decrease problem, the most efficient solution is addition, not subtraction. Taking a snapshot of a change that corresponds with a subtraction is problematic, because it is not possible to simultaneously picture a yet unchanged start set (the minuend), the ongoing change, i.e., the subtraction set (which was part of the start set) and the unknown difference set in a static representation. A limitation of the present study is that the increase-decrease dimension cannot be separated from the unknown dimension (first position versus

last position). This limitation is, however, a result of our choice to stay close to the original problems that are used in education. Direction of the change and location of the unknown are inherently interconnected in these types of pictorial problems. Another characteristic of these decrease problems is that the curriculum developers used a minus sign in the accompanying schematic number sentence. This minus contradicts the addition, which is required to correctly solve the problem, and can, therefore, be confusing. Both issues should be addressed in future studies.

One could conclude that decrease problems in the form of a static picture are ill chosen by developers of mathematics curricula. However, this study was not a teaching experiment. We do not know whether, and if yes, how these problems are taught by teachers in class, nor do we know experiments that were aimed at training to solve these problems. Our findings highlight that educators must be aware of the difficulties imposed on the children when using these types of problems. Further research is needed to investigate whether these pictorial decrease problems are harmful or, instead, facilitate deeper learning of mathematical relations.

In the Introduction, we pointed out that there are also pictorial problems, similar to the case of mathematical word problems, in which different semantic structures like change, combine and compare are distinguished. It is possible that, just like with different word problems formats, other pictorial problems aside from the ones used in this study, are easier or more difficult to solve. For example, pictures with a combine structure, i.e., a structure in which sets are connected to each other in a part-whole relation, are possibly better to comprehend than statically represented decrease problems in the change format. Future research should investigate the relative merits of pictorial problems with these different underlying semantic structures. Also, pictorial change type problems, like the ones studied in the present experiment, could perhaps still be more helpful when interactive computer animations are used to present them dynamically instead of a static picture - see for example a study by Scheiter, Gerjets, and Schuh (2010) regarding high school students and how they solve algebra word problems. Given our findings, adding auditory support in educational software could be helpful as well. To our knowledge, no systematic research has been carried out on pictorial representations of simple addition and subtraction problems. We think that in this area foundational research must be conducted to help build a theory of the effects of pictorial problems and the way they can be used optimally in exercises and teaching, similarly to what other scientists have done in the area of arithmetic word problem solving (Carpenter & Moser, 1984; De Corte et al., 1985; Thevenot, Devidal, Barrouillet, & Fayol, 2007). The present study's findings comprise a stepping-stone towards this direction.

4.1 Conclusions

We showed that children perform worse in decrease problems represented with static pictures compared to corresponding increase problems. This was because they imposed more cognitive load. Performance improved when we decreased the assumed high cognitive load by supporting the pictorial decrease problems with auditory, i.e., spoken, information. Taken together, our findings suggest that developers of mathematics curricula should be aware of what is achieved by using static pictures for decrease problem situations, which suggest a subtraction but are in fact requiring addition instead.

Acknowledgements We thank Anniek Gerrits and Sabina van den Bor for collecting the data for the experiment. We also thank the participating school and children.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest. The research involved human participants (children), who participated in the study after we had received informed consent.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Hoboken: Wiley.
- Baddeley, A. D. (2009). Working memory. *Science*, *255*, 556–559.
- Berends, I. E., & van Lieshout, E. C. D. M. (2009). The effect of illustrations in arithmetic problem-solving: Effects of increased cognitive load. *Learning and Instruction*, *19*, 345–353. <https://doi.org/10.1016/j.learninstruc.2008.06.012>
- Campbell, J. I. D., & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General*, *130*, 299–315. <https://doi.org/10.1037/0096-3445.130.2.299>
- Campbell, J. I. D., Fuchs-Lacelle, S., & Phenix, T. L. (2006). Identical elements model of arithmetic memory: Extension to addition and subtraction. *Memory & Cognition*, *34*, 633–647. <https://doi.org/10.3758/BF03193585>
- Carpenter, T. P., & Moser, J. M. (1984). *Journal for Research in Mathematics Education*, *15*, 179–202.
- Carpenter, T. P., Moser, J. M., & Bebout, H. C. (1988). *Journal for Research in Mathematics Education*, *19*, 345–357.
- De Corte, E., & Verschaffel, L. (1981). Children's solution processes in elementary arithmetic problems: Analysis and improvement. *Journal of Educational Psychology*, *73*, 765–779.
- De Corte, E., Verschaffel, L., & De Win, L. (1985). Influence of rewording verbal problems on children's problem representations and solutions. *Journal of Educational Psychology*, *77*, 460–470.
- Dewolf, T., Van Dooren, W., Hermens, F., & Verschaffel, L. (2015). Do students attend to representational illustrations of non-standard mathematical word problems, and, if so, how helpful are they? *Instructional Science*, *43*, 147–171. <https://doi.org/10.1007/s11251-014-9332-7>
- Dewolf, T., Van Dooren, W., & Verschaffel, L. (2016). Can visual aids in representational illustrations help pupils to solve mathematical word problems more realistically? *European Journal of Psychology of Education*. <https://doi.org/10.1007/s10212-016-0308-7>
- Fuchs, L. S., Fuchs, D., Compton, D. L., Hamlett, C. L., & Wang, A. Y. (2015). Is word-problem solving a form of text comprehension? *Scientific Studies of Reading*, *19*, 204–223. <https://doi.org/10.1080/10888438.2015.1005745>
- Garson, G. D. (2013). *Generalized linear models/generalized estimating equations*. Asheboro: Statistical Publishing Associates.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446. <https://doi.org/10.1016/j.jml.2007.11.007>
- Janssen, J., & Kraemer, J. M. (2002). *Rekenen-wiskunde 2002* [Arithmetic and Mathematics 2002]. Arnhem: Cito.
- Jenks, K. M., de Moor, J., & van Lieshout, E. C. D. M. (2009). Arithmetic difficulties in children with cerebral palsy are related to executive function and working memory. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *50*, 824–833. <https://doi.org/10.1111/j.1469-7610.2008.02031.x>
- Kamii, C., Lewis, B. A., & Kirkland, L. D. (2001). Fluency in subtraction compared with addition. *The Journal of Mathematical Behavior*, *20*, 33–42. [https://doi.org/10.1016/S0732-3123\(01\)00060-8](https://doi.org/10.1016/S0732-3123(01)00060-8)
- Kintsch, W. (1986). Learning from text. *Cognition and Instruction*, *3*, 87–108.
- Leikin, R., Leikin, M., Waisman, I., & Shaul, S. (2013). Effect of the presence of external representations on accuracy and reaction time in solving mathematical double-choice problems by students of different levels of instruction. *International Journal of Science and Mathematics Education*, *11*, 1049–1066.

- Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358–368. <https://doi.org/10.1037/0022-0663.91.2.358>
- Pap, W. (2002). *Pluspunt 2* [Plus Point 2]. 's Hertogenbosch, The Netherlands: Malmberg.
- Passolunghi, M. C., Vercelloni, B., & Schadee, H. (2007). The precursors of mathematics learning: Working memory, phonological ability and numerical competence. *Cognitive Development, 22*, 165–184. <https://doi.org/10.1016/j.cogdev.2006.09.001>
- Rasmussen, C., & Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of Experimental Child Psychology, 91*, 137–157. <https://doi.org/10.1016/j.jecp.2005.01.004>
- Scheiter, K., Gerjets, P., & Schuh, J. (2010). The acquisition of problem-solving skills in mathematics: How animations can aid understanding of structural problem features and solution procedures. *Instructional Science, 38*, 487–502. <https://doi.org/10.1007/s11251-009-9114-9>
- Stellingwerf, B. P., & van Lieshout, E. C. D. M. (1999). Manipulatives and number sentences in computer aided arithmetic word problem solving. *Instructional Science, 27*, 459–476.
- Stylianou, D. A., & Silver, E. A. (2004). The role of visual representations in advanced mathematical problem solving: An examination of expert–novice similarities and differences. *Mathematical Thinking and Learning, 6*, 353–387.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295–312. [https://doi.org/10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5)
- Sweller, J., van Merriënboer, J. G., & Paas, F. G. W. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296.
- Thevenot, C. (2010). Arithmetic word problem solving: Evidence for the construction of a mental model. *Acta Psychologica, 133*, 90–95. <https://doi.org/10.1016/j.actpsy.2009.10.004>
- Thevenot, C., Devidal, M., Barrouillet, P., & Fayol, M. (2007). Why does placing the question before an arithmetic word problem improve performance? A situation model account. *The Quarterly Journal of Experimental Psychology, 60*, 43–56. <https://doi.org/10.1080/17470210600587927>
- Verschaffel, L., De Corte, E., & Lasure, S. (1994). Realistic considerations in mathematical modeling of school arithmetic word problems. *Learning and Instruction, 4*, 273–294.
- Xenidou-Dervou, I., Molenaar, D., Ansari, D., van der Schoot, M., & van Lieshout, E. C. D. M. (2017). Nonsymbolic and symbolic magnitude comparison skills as longitudinal predictors of mathematical achievement. *Learning and Instruction, 50*, 1–13. <https://doi.org/10.1016/j.learninstruc.2016.11.001>