Young children's inhibition of keyword heuristic in solving arithmetic word problems

Hei Yu Shum[1], Winnie Wai Lan Chan[1]*

Arithmetic word problems are particularly difficult to solve when they are presented in inconsistent language, where the keyword (e.g. "*more than*") does not match with the correct arithmetic operation (e.g. subtraction). To solve these problems, one should abstain from deriving the arithmetic operation directly from the keyword (keyword heuristic). Using a negative priming task, this study examined the need of inhibiting the keyword heuristic among 9- to 10-year-olds and compared the inhibition efficiency of children with differential problem-solving abilities. Results showed a negative priming effect, suggesting that the keyword heuristic had been inhibited when solving the inconsistent-language problem. Importantly, there was no difference in inhibition efficiency between the above-average and below-average problem solvers. Hence, at least for elementary-grade children, what differentiates their problem-solving abilities may not be related with their efficiency in inhibiting a misleading heuristic.

Keywords: arithmetic word problems, inhibitory control, inconsistent language

INTRODUCTION

Arithmetic word problem solving can be a challenge for students from elementary school through adulthood (*1*). When presented in inconsistent language (in which the relational term does not match with the correct arithmetic operation), the problems are even harder as the commonly used keyword heuristic (translating keywords [e.g. "*more than*" to arithmetic operations [e.g. addition]) becomes misleading and may need to be inhibited (*2*). However, it remains unexplored whether inhibitory control differentially affects young children with worse problem-solving ability. In this research, we adopted the negative priming task (*2*) to examine the need of inhibiting heuristic in young children (i.e. fourth graders) and compare the inhibition efficiency of problem-solvers with different abilities. This may help us gain better insights into how best to support children in solving inconsistent-language problems.

Consistency Hypothesis

Riley, Greeno and Heller (*3*) identified four types of problems: Change and equalise problems involve actions that lead to changes in quantity, whereas combine and compare problems describe static relations between two quantities. Compare problems are the most cognitively demanding, requiring the construction of complex schemas (*3*). Lewis and Mayer's (*4*) consistency hypothesis suggested difficulties may arise from the inconsistent-language format (e.g. in "*David has 10 pens. **He has 5 fewer pens than Mary [unknown set].** How many pens does Mary have?*"), as the unknown set is the object of the relational statement and the relational term ("*fewer than*") contradicts the correct operation (addition: 10 + 5). By contrast, the unknown set in a consistent-language problem is the subject (e.g. "*David has 10 pens. **Mary [unknown set] has 5 more pens than David.** How many pens does Mary have?*"). The relational term ("*more than*") matches with the correct arithmetic operation (addition: 10 + 5) and provides useful hint about which operation to use. This hypothesis has been supported by the findings that more errors were committed by college students (*4*) and more time was required by children to read and solve inconsistent-language problems than consistent-language ones (*5*).

Inhibition of Keyword Heuristic

The neo-Piagetian models suggested one may need to inhibit overlearned heuristics in cases when they are not useful and activate the more cognitively demanding analytical strategies (*6*, *7*). Solving inconsistent-language problems may require inhibiting irrelevant heuristics (*2*). One common heuristic reinforced in mathematics classrooms for word problems is the keyword or "*add if more, subtract if less*" heuristic (directly adopting the arithmetic operation [e.g. addition/subtraction] hinted by the keywords [e.g. "*more than/less than*"]) (*8*). By contrast, analytical strategy takes a meaningful problem model approach in which one analyses the problem statement carefully to translate it into a mental model of the situation (*9*). Using the heuristic arbitrarily in inconsistent-language problems may lead to *reversal errors* (adopting the inverse of the correct operation) which are common among school-aged and college students (*8*, *10*).

To test whether inhibition of the heuristic is required, Lubin et al. (*2*) designed a negative priming (NP) paradigm. The NP paradigm has been widely adopted to examine inhibitory control in situations involving conflicting information (*11*, *12*). In Lubin et al.'s (*2*) NP paradigm, participants saw problem pairs in a prime-probe sequence. In the test trials, an inconsistent-language problem (where the keyword heuristic should be inhibited) appeared first as the prime, followed by a consistent-language probe (where the heuristic should be re-activated). In the control trials, a neutral problem (where the heuristic was irrelevant) was presented as the prime, followed by a consistent-language probe. The principle is that activation of a previously inhibited heuristic requires more cognitive efforts than activation of the same heuristic without previous inhibition (*13*). Hence, if inhibition was required, there would be an NP effect shown by longer response times or higher error rates for solving the probes in the test trials than in the control trials (*2*). Indeed, a significant NP effect in response times (despite no group differences) was found in sixth-graders, adolescents and young adults, suggesting an executive cost to re-activate the keyword heuristic which had previously been inhibited when solving the inconsistent-language problem (*2*, *14*). Since solution procedures needed to be changed between primes and probes, it is unlikely that NP effects reflect only the cost of moving to a different problem type. Moreover, the NP effect has only been tested among 11- to 12-year-olds, adolescents and young adults (*2*, *14*), it remains unexplored whether it exists in younger children as well. In other words, it is unclear whether the keyword heuristic is overlearned and dominant among younger children to the extent that it warrants inhibition in the case of inconsistent-

[1]Department of Psychology, The University of Hong Kong, Hong Kong SAR, China.
*Corresponding author. Email: wlwinnie@hku.hk (W.W.L.C)

language problem-solving.

Inhibition and Word Problem Solving

The relationship between inhibition control and arithmetic word problem solving has been examined in previous studies. For example, executive function components (including response inhibition, shifting and working memory) have been found to affect third-graders' arithmetic word problem solving skills (*15*); and fifth-graders with greater executive function capacities (i.e. perseveration, inhibition efficiency, working memory span, etc.) performed better on algebra word problems (*16*). Moreover, a higher number of irrelevant answers to word problems were produced by 10-year-olds with Attention Deficit/Hyperactivity Disorder (who had impairment in inhibitory control) compared to typically developing children (*17*). It has been suggested that inhibitory control is needed to suppress the dominant but inefficient strategies for the execution of new but more efficient strategies (*18*) when the problem text involves irrelevant information that adds demands on data selection (*19*). However, research focusing specifically on inhibitory control and performance on inconsistent-language word problems has been scarce (*2, 14*). One exception was the study by Lubin et al. (*14*), which adopted the NP paradigm and found smaller magnitudes of the NP effect among the math experts (maths undergraduates) than the non-experts (non-math undergraduates). This suggested that the math experts were more efficient in inhibiting the misleading keyword heuristic. More studies are needed to investigate whether children who are poor at solving inconsistent language problems are actually less efficient in inhibiting the keyword heuristic.

The Present Research

This study aimed to address two questions: (1) Do young children need to inhibit the keyword heuristic when solving inconsistent-language word problems? (2) If yes, do good and poor word-problem solvers differ in their inhibition efficiency? To address the first question, we examined the need of inhibition among 9- to 10-year-olds. Note that children younger than 9 years old may not reliably exhibit an NP effect as their inhibitory mechanism is still developing (*20*). Based on Lubin et al. (*2*), we hypothesised that young children would need to inhibit the heuristic when solving inconsistent-language problems, which would be indicated by an NP effect representing the extra cost (indicated by lower problem solving efficiency) incurred in re-activating the heuristic inhibited in the inconsistent-language prime to solve the consistent-language probe in the test trials than in the control trials. For the second question, based on Lubin et al.'s study (*14*) showing better inhibition efficiency among maths experts, we hypothesised that children who were weaker problem-solvers would be less efficient in inhibiting the keyword heuristic (indicated by larger magnitudes of the NP effect). Note that the children in this study were Cantonese-speaking, thus the word problems were presented in Chinese. The linguistic structure of problem presentation in Chinese is similar to that in English. Take this as example: "*Ben has 8 apples. He has 5 more apples than Jenny.*" This is expressed in Chinese as ''小明 [Ben] 有 [has] 8 [numeral] 個 [quantifier] 蘋果 [apples], 比 [compare with] 小珍 [Jenny] 多 [more] 5 [numeral] 個 [quantifier]''. The results of this study will help us better understand whether inhibition of irrelevant strategy poses a challenge to some Chinese children when solving inconsistent-language problems, thereby offering insights into how best to facilitate children in solving such problems.

RESULTS

Seventy-seven Cantonese-speaking fourth graders (mean age ($\pm SD$) = 9.72 $\pm$ .30 years; 34 boys and 43 girls) participated in the study. All participants completed a word problem test for differentiating word problem solving ability, followed by a negative priming task to test inhibition.

For the word problem test, one point was given for each correct number sentence. The test score was the total number of correct number sentences for the target problems. The mean was 5.70 (*SD* = 3.20). Two groups were divided using mean-split method, namely the above-average (n = 34; 21 males and 13 females; *M* = 8.41, *SD* = 2.11) and the below-average problem-solvers (n = 36; 10 males and 26 females; *M* = 3.14, *SD* = 1.48).

For the negative priming task, we excluded children showing inattentive behaviour (n = 7) reported by the experimenters, including pressing response buttons randomly, chatting with classmates, being distracted by the surroundings and looking around. Analyses were based on 70 children (mean age: 9.67 $\pm$ .30 years; 31 males and 39 females). Only trials in which both the primes and probes were correct were included for analyses of response times. Outliers were defined as response times beyond 2 SDs from the individual means in the primes and probes of the test and control conditions respectively (*2*). After removing the outliers (2.4% of the trials), accuracy rates and response times were averaged for each child to obtain separate scores for the primes and probes of the test and control conditions respectively. The inverse efficiency score which represented the efficiency of problem solving was computed by dividing the mean response time by the mean accuracy rate of each participant (*21*). The higher the inverse efficiency score, the less efficient the performance would be. Separate analyses of variance (ANOVAs) were conducted on the inverse efficiency scores for the primes and probes, with condition (test/control) as the within-participants factor and ability level (above-average/below-average) as the between-participants factor (see **Table 1** for a summary).

Concerning the primes, a significant main effect of condition was found, $F(1, 68) = 7.43$, $p = .008$, $\eta^2 = .10$. The children's performance was less efficient in the test condition (i.e. inconsistent-language problems) ($M = 164.02$, $SD = 184.33$) than in the control condition (i.e. neutral problems) ($M = 101.96$, $SD = 67.35$). However, no significant main effect of ability level was found, $F(1, 68) = 3.41$, $p = .07$, $\eta^2 = .05$. There was no significant interaction effect between ability level and condition as well, $F(1, 68) = .55$, $p = .46$, $\eta^2 = .01$. The lack of group difference (insignificant main effect of ability level) was probably due to the sole use of one-step problems in the NP task which were supposed to have been learned in first grade and were thus relatively easy to many fourth graders. Hence, the task might not be sensitive enough to differentiate the problem solving abilities among fourth graders. However, it should be noted that in order to yield a reliable score for inverse inefficiency, accuracy should not be too low because response times of the incorrect trials would be excluded. This was the reason why we included simple problems in the NP task to ensure a reasonable accuracy rate and used the word problem test which included multiple-step problems for differentiating the two groups of problem solvers.

Regarding the probes, we hypothesised that there would be an NP effect indicated by a main effect of condition, in which the performance on test condition should be less efficient (i.e. higher inverse efficiency score) than the control condition. Indeed, a significant main effect of condition was found, $F(1, 68) = 6.53$, $p = .013$, $\eta^2 = .09$. The children performed less efficiently in the test

Table 1. Descriptive statistics.

	Accuracy rates (%)		Response times (ms)		Inverse Efficiency (ms)		NP scores (ms)
	Prime	Probe	Prime	Probe	Prime	Probe	
	Test Control	Test Control	Test Control	Test Control	Test Control	Test Control	
All participants *M* (*SD*)	71.43 (26.93) 88.75 (17.31)	85.18 (17.84) 88.21 (14.25)	8,206.27 (3,089.99) 8,263.32 (2,195.88)	8,766.89 (2,702.60) 8,193.91 (2,475.45)	164.02 (184.33) 101.96 (67.35)	109.51 (54.99) 93.47 (26.91)	16.04 (51.96)
Above-average group *M* (*SD*)	83.46 (20.35) 92.28 (13.06)	91.54 (13.67) 90.44 (12.71)	7,999.91 (3,363.51) 8,620.59 (2,510.16)	9,141.09 (2,237.56) 8,457.98 (2,323.16)	132.64 (190.44) 87.81 (26.13)	104.37 (40.28) 95.23 (28.72)	9.14 (39.45)
Below-average group *M* (*SD*)	60.07 (27.68) 85.42 (20.16)	79.17 (19.36) 86.11 (15.46)	8,424.76 (2,803.00) 7,885.04 (1,764.27)	8,413.47 (3,068.27) 7,944.51 (2,619.21)	193.65 (175.88) 115.32 (89.00)	114.37 (66.19) 91.81 (25.38)	22.56 (61.36)

condition ($M = 109.51$, $SD = 54.99$) than in the control condition ($M = 93.47$, $SD = 26.91$). Yet, there was no significant main effect of ability level, $F(1, 68) = .16$, $p = .69$, $\eta^2 = .002$. No significant interaction effect between ability level and condition was found, $F(1, 68) = 1.17$, $p = .28$, $\eta^2 = .02$.

Given the significant NP effect, we further compared the magnitudes of the effects between the two ability groups. An NP score was obtained for each participant to represent the cognitive cost of inhibition by subtracting the individual average inverse efficiency score on the probe in the control condition from the test condition. An independent-samples *t*-test showed no significant difference in the NP scores between the above-average group ($M = 9.14$, $SD = 39.45$) and the below-average group ($M = 22.56$, $SD = 61.36$), $t(68) = 1.08$, $p = .28$, $d = .26$. This suggested that both groups did not differ in the inhibition efficiency.

Further analysis was conducted to examine possible gender difference in the magnitude of the NP effect. An independent-samples *t*-test showed no significant difference between boys ($M = 17.59$, $SD = 62.50$) and girls ($M = 14.88$, $SD = 43.23$), $t(68) = .21$, $p = .83$, $d = .05$. This indicated that no gender difference was found in the inhibition efficiency.

DISCUSSION

Inhibition of Keyword Heuristic

The present research aimed to explore whether inhibition of keyword heuristic is required when young children solve inconsistent-language word problems. We examined whether 1) inhibitory control was needed in solving inconsistent-language problems among fourth graders and 2) whether solvers with differential abilities showed different NP magnitudes. Regarding the first question, an NP effect was found which reflected an executive cost of activating the keyword heuristic previously inhibited to solve the inconsistent-language prime. The finding of this study served to extend the results of Lubin et al. (*2, 14*) to younger children (9- to 10-year-olds) solving Chinese word problems, as this is the first research focusing on solving inconsistent language arithmetic word problems in Chinese. From daily life experiences, children learn to associate "*more than*" or "*less than*" with addition or subtraction respectively (*2*). This straightforward heuristic is readily accessible and heavily relied on (*9*). Indeed, direct translation of the relational term is a common error when solving inconsistent

language problems among Cantonese-speaking children (*22*). However, when language is inconsistent, extra cognitive efforts may be required to inhibit this heuristic from interfering the execution of other strategies.

Indeed, our study added an important cognitive step to the problem representation model suggested by Lewis and Mayer (*4*) (see **Fig. 1**). The original model describes the steps involved in solving an inconsistent language word problem which consists of three sentences with the second sentence containing the misleading relational term.
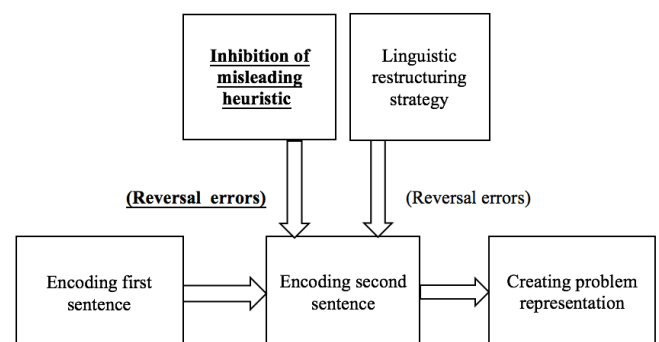
Fig. 1. Modified model of problem representation phase (*2, 4*). (Modified parts are in bold and underlined).

According to the model, the problem is encoded sentence by sentence to create a problem presentation. When encoding the second relational sentence, one should mentally restructure the statement and reverse the arithmetic operation primed by the relational term. This step is known as the linguistic restructuring strategy (*4*). The present findings suggest that inhibition of heuristic may occur in the process of encoding the relational sentence. Take this problem as an example: "*Casey has 10 apples. She has 5 apples more than Sherry. How many apples does Sherry have?*" During problem representation, the solver first encodes the first sentence (*Casey has 10 apples*) and instantiates the assignment schema as *Casey = 10 apples* (*4*). When encoding the second sentence (*She has 5 apples more than Sherry*), the solver locates the subject (*She, which refers to Casey*), the number (*5*), the relational term (*more than*) and the name *Sherry* (*4*). This is where inhibitory control is likely to be at play, as the "*add if more, subtract if less*"

heuristic may be activated. The relational term (*more than*) may automatically prompt the activation of the corresponding arithmetic operation (addition). If the solver fails to inhibit the heuristic (*add if more*), he/she may directly create a problem representation according to the current schema (i.e. *Sherry's apples = 10 + 5*), leading to reversal errors. If, however, the solver notices that the subject in the second sentence (*She*) matches that in the first sentence (*Casey*), the solver may realise that the keyword heuristic (*add if more*) is actually misleading and thus should be inhibited. During this time, the linguistic restructuring strategy can be applied, in which the order of the two protagonists in the second sentence (*She/Casey and Sherry*) are reversed, so that *Sherry has 5 apples more than Casey*. Then the relational term is reversed accordingly, i.e. *more than* is changed to *fewer than*. The process of linguistic restructuring may fail due to high working memory demands. This is another critical point where reversal errors may occur. After successful restructuring, a relational schema is instantiated and an equation (number sentence) is created (*Sherry's apples = 10 - 5*), followed by problem solution which involves planning, monitoring and execution of arithmetic computations (*4*). Thus, the modified model suggests that inhibition of heuristic may be an essential step involved in encoding the relational sentence in the problem, so as to allow the execution of the linguistic restructure strategy.

The second question asked in this study was whether children with different performance in word problem solving showed different NP magnitudes. To our knowledge, this was the first study to examine the relationship between performance in arithmetic word problem solving and inhibition efficiency in Chinese children at fourth grade. Interestingly, children who were above-average and below-average problem solvers did not differ in their inhibition efficiency. This suggests that the differential performance in word problem solving may not be related to inhibition efficiency at this age. This might be due to the development of inhibitory mechanisms. NP effects may reflect *automatic inhibition* processes (ignoring stimuli) which rely on posterior sensory parts that become mature during early childhood or *intentional inhibition* processes (resisting a misleading strategy) which rely on prefrontal cortex that is not yet fully mature until late adolescence (*23, 24*). Solving inconsistent-language problems may involve both because one may need to ignore the relational term and the heuristic (*2*). Perhaps 9- to 10-year-olds are still developing *intentional inhibition control*, so individual differences in inhibitory efficiency may not be salient enough for any demarcation to emerge. Other cognitive skills such as working memory (*22*), language ability (*25*), attention (*26*), and nonverbal intelligence (*26*) may be more closely linked to the differential performance.

Concerns on Negative Priming Paradigm
Negative priming has been linked to *selective inhibition*, which attributes the disruption in the response of the probe to the inhibition of the target in the prime (*11*). However, the *episodic retrieval account* suggests that the probe stimulus serves as a retrieval cue for prior processing of the same stimulus on the prime, in which a "*do not respond tag*" previously associated with the prime became associated with the probe, resulting in delayed response in the probe (*11*). In our study, it is unlikely that the NP effect was due to episodic retrieval, because different stimuli (i.e. objects and protagonists) were used for the primes and probes. One may also argue that the NP effects only represented the cognitive cost when one shifted from one strategy to another between the prime and the probe problems (*2*), rather than the cost of activating a

strategy which was previously inhibited. Nevertheless, the children actually needed to adopt different strategies for the prime and probe problems in both conditions (*2*). Hence it is likely that the NP effects reveal more than just shifting cost.

Implications
Our findings provide useful directions for helping children tackle inconsistent-language problems. First, since the keyword heuristic appears to be a default strategy and can sometimes be misleading, children need to be taught why and when the heuristic does not work in some cases, and what to look for in a word problem to determine whether the heuristic should be applied or inhibited. Explicit instruction on the linguistic patterns such as the reversibility of relational terms may help, as the awareness of the linguistic expressions is the key to solving word problems (*1*). Second, metacognitive inhibition training could be provided. Teaching the significance of inhibiting inappropriate strategies and adopting the correct ones has shown to effectively improve logical reasoning (*27, 28*). The inhibition process may be visualised using animations or picture cards, so that children may learn to suppress the irrelevant strategy when it pops up.

Limitations and Future Work
First, due to the nature of NP paradigm, our results could only be interpreted to suggest the inhibition of a strategy without providing much information on the nature of the strategy itself (*14*). Future research could employ additional methods such as verbal reports to examine the nature of alternative strategies used. Second, the NP results were only based on trials that were correct on both primes and probes, so such findings may only be relevant to the successfully solved problems. As for the inconsistent-language problems that the students failed to solve, it remains unclear whether ineffective inhibition of the heuristic is the underlying factor. Third, two ability groups were divided using the mean-split method and this might have limited the variability of performance. Further studies may recruit a larger representative pool to obtain better sampling of good and poor problem solvers. Lastly, this study lacked behavioural tests such as measures of executive control and general inhibition ability. Future study may include such tests to provide more elaborated explanations about the role of inhibition in word problem solving.

Conclusions
The present research examined the inhibition of keyword heuristic when solving inconsistent-language arithmetic word problems among Chinese fourth graders. The findings suggested that cognitive efforts are needed in inhibiting the heuristic during the problem representation process, although there may not be individual differences in inhibition efficiencies at this age. Instructional supports on proper use of the heuristic and inhibitory control may be potential directions to help children better tackle such problems.

MATERIALS AND METHODS
Experimental Design
We used the negative priming paradigm (*2*) to examine the necessity of inhibition of keyword heuristic when solving inconsistent-language word problems among 9- to 10-year-olds. To examine whether good and poor word-problem solvers differ in their inhibition efficiency, we further compared their magnitudes of NP effect.

Participants

A total of 77 Chinese fourth graders (mean age ($\pm SD$) = 9.72 $\pm$ .30 years; 34 boys and 43 girls) were recruited from two primary schools in Hong Kong. An *a priori* power analysis, using the statistical program G*Power 3.1.9.2 (*29*), indicated that a sample size of 54 children would be adequate to achieve 80% power with a medium effect size. Informed consent was obtained from the children and their parents after the nature and possible consequences of the studies were explained.

Negative Priming Task

The task adopted the negative priming paradigm (*2*). All word problems comprised three statements that described quantity of objects or relations of the quantities. The word problem types included inconsistent-language, consistent-language and neutral problems (presented in 35-point white Courier New font on a black background) (see **Fig. 2**).
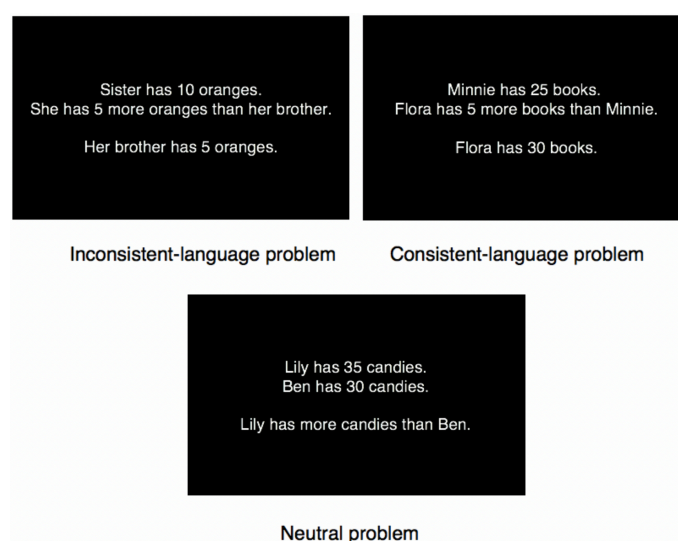

Fig. 2. **Samples of the negative priming task.** Problems were presented in Chinese.

For inconsistent- and consistent-language problems, the first two statements described a) the quantity of objects possessed by a character and b) the difference in quantity of objects between that character and another character, described by a relational term. Half of the trials contained the relational term "*more than*" and the other half contained "*less than*". For neutral problems, each of the first two statements described the quantity of objects possessed by a character. In all problems, the last sentence was a probable statement that may be derived from the first two sentences regarding the quantity of objects possessed by one of the characters (in inconsistent- or consistent-language problems) or a relational difference in the quantity of objects possessed by the two characters (in neutral problems). Children were asked to decide whether the proposed statement was correct. Only basic adding or subtracting of numbers divisible by 5 (e.g. 5, 10, 25, 40) was required to minimise the difficulty of calculation. Each problem was presented with a different content with different names of characters and objects to avoid retention of information.

Word Problem Test

The test was designed based on Grade 4 mathematics supplementary exercises to measure ability in solving inconsistent-language word problems. The 15-minute test comprised 20 randomised arithmetic word problems in Chinese, including 15 inconsistent-language problems (2 one-step, 7 two-step and 6 multiple-step) as target problems and 5 consistent-language filler problems (2 two-step and 3 multiple-step) to avoid practice effect. In the one-step problem, the solution required only one arithmetic operation (e.g. "*There were some egg tarts in the bakery. After 176 egg tarts were sold, there are now 34 left. How many egg tarts were there in the bakery originally? Answer: 176 + 34*"). Two arithmetic operations were required in the two-step problems (e.g. "*Lily's cup has 390 mL of orange juice. After drinking 180 mL of orange juice in her own cup and then adding 50 mL of orange juice into the cup, Janice now has the same amount of orange juice as Lily. How much orange juice did Janice's cup have originally? Answer: 390 + 180 − 50*"). Multiple-step problems required three or more arithmetic operations (e.g. "*There are 56 pears in the garden, which are 10 more than oranges. There are 8 fewer apples than oranges. What is the difference between pears and apples? Answer: 56 - (56 - 10 - 8)*"). All problems involved addition and/or subtraction only. Four problem types were involved (*3*), consisting of 6 compare problems, 4 combine problems, 6 change problems and 4 equalise problems. To rule out calculation skills, only the number sentence was required as answer for each question.

Procedure

Negative Priming Task

Each child was tested individually with an experimenter on a laptop. Stimuli were shown on E-prime 2.0 (Psychological Software Tools, Inc., Pittsburgh, PA). Each trial contained a word problem pair in a prime-probe sequence. In the test condition, an inconsistent-language problem was first presented as the prime, followed by a consistent-language probe (**Fig. 3**).
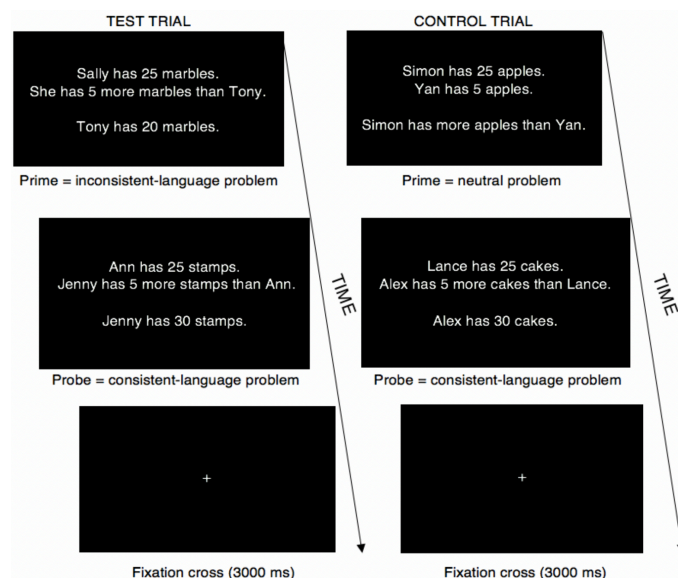

Fig. 3. **Sample procedures of the negative priming task.** A fixation cross was displayed between each trial for 3000 ms to limit retention of previous information. Problems were presented in Chinese.

In the test trials, the keyword strategy that participants had to inhibit to solve the prime became the useful strategy for the probe. To solve the prime in **Fig. 3**, children would need to inhibit performing addition (25 + 5; the misleading strategy hinted) in order to elicit subtraction (25 − 5). The consistent-language probe then immediately followed, and children would need to adopt the keyword strategy which was supposed to have been inhibited

previously in the prime. To solve the probe in **Fig. 3**, children were required to perform addition (25 + 5; correct strategy hinted by the keyword). In the control condition, the strategy for solving the prime (a neutral problem) was irrelevant to the strategy for solving the probe (a consistent-language problem), and the heuristic was thus neither activated nor inhibited. In the neutral prime, children had to compare the number of objects that each person possessed without involving arithmetic operations (see **Fig. 3**). The prime was immediately followed by a consistent-language probe.

Children decided whether the proposed answer was correct by pressing the yellow ("f") button for correct and the blue ("j") button for incorrect. Children first completed three practice trials with feedback. Then, children performed 16 randomised experimental trials including 8 test and 8 control trials without feedback (no more than two test or control trials would appear consecutively). Each trial was shown until one of the two buttons was pressed. A fixation cross was displayed in the middle of the screen between trials for 3000 ms. Response times were the duration from the onset of the problem to the pressing of the key.

Word Problem Test

Children in small groups were verbally instructed to write down the number sentence without performing any calculation and complete as many problems as possible within 15 minutes. An example was first explained by the experimenter. Children completed all questions on their own.

REFERENCES

1. L. Verschaffel, Using retelling data to study elementary school children's representations and solutions of compare problems. *J. Res. Math Educ.* **25**, 141-165 (1994).
2. A. Lubin, J. Vidal, C. Lanoë, O. Houdé, G. Borst, Inhibitory control is needed for the resolution of arithmetic word problems: A developmental negative priming study. *J. Educ. Psychol.* **105**, 701-708 (2013).
3. M. S. Riley, J. G. Greeno, J. H. Heller, *The Development of Mathematical Thinking*, H. P. Ginsberg, ed. (Academic Press, 1983), pp. 153-196.
4. A. B. Lewis, R. E. Mayer, Students' miscomprehension of relational statements in arithmetic word problems. *J. Educ. Psychol.* **79**, 363-371 (1987).
5. L. Verschaffel, E. D. Corte, A. Pauwels, Solving compare problems: An eye movement test of Lewis and Mayer's consistency hypothesis. *J. Educ. Psychol.* **84**, 85-94 (1992).
6. O. Houdé, First insights on "neuropedagogy of reasoning". *Thinking & Reasoning* **13**, 81-89 (2007).
7. R. Siegler, *Emerging Minds: The Process of Change in Children's Thinking* (Oxford University Press, 1996).
8. M. Hegarty, R. E. Mayer, C. E. Green, Comprehension of arithmetic word problems: Evidence from students' eye fixations. *J. Educ. Psychol.* **84**, 76-84 (1992).
9. M. Hegarty, R. E. Mayer, C. A. Monk, Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *J. Educ. Psychol.* **87**, 18-32 (1995).
10. A. Lewis, Training students to represent arithmetic word problems. *J. Educ. Psychol.* **81**, 521-531 (1989).
11. G. Borst, S. Moutier, O. Houdé, *New Approaches in Reasoning Research*, W. D. Neys, M. Osman, eds. (Taylor and Francis, 2013), pp. 34-50.
12. S. P. Tipper, M. Cranston, Selective attention and priming: Inhibitory and facilitatory effects of ignored primes. *Q. J. Exp. Psychol. A.* **37**, 591-611 (1985).
13. S. P. Tipper, The negative priming effect: Inhibitory priming by ignored objects. *Q. J. Exp. Psychol. A.* **37**, 571-590 (1985).
14. A. Lubin, S. Rossi, C. Lanoë, J. Vidal, O. Houdé, G. Borst, Expertise, inhibitory control and arithmetic word problems: A negative priming study in mathematics experts. *Learn. Instr.* **45**, 40-48 (2016).
15. P. Viterbori, L. Traverso, M. C. Usai, The role of executive function in arithmetic problem-solving processes: A study of third graders. *J. Cogn. Dev.* **18**, 595-616 (2017).
16. K. Lee, S. F. Ng, E. L. Ng, Z. Y. Lim, Working memory and literacy as predictors of performance on algebraic word problems. *J. Exp. Child Psychol.* **89**, 140-158 (2004).
17. S. Sabagh-Sabbagh, D. A. Pineda, Cognitive inhibitory control and arithmetic word problem solving in children with attention deficit/ hyperactivity disorder: A pilot study. *Univ. Psychol.* **9**, 761-772 (2010).
18. K. H. Khng, K. Lee, Inhibiting interference from prior knowledge: Arithmetic intrusions in algebra word problem solving. *Learn. Individ. Differ.* **19**, 262-268 (2009).
19. M. C. Passolunghi, L. S. Siegel, Short-term memory, working memory, and inhibitory control in children with difficulties in arithmetic problem solving. *J. Exp. Child Psychol.* **80**, 44-57 (2001).
20. S. P. Tipper, T. A. Bourque, S. H. Anderson, J. C. Brehaut, Mechanisms of attention: A developmental study. *J. Exp. Child Psychol.* **48**, 353-378 (1989).
21. W. W. L. Chan, T. K. Au, J. Tang, Developmental dyscalculia and low numeracy in Chinese children. *Res. Dev. Disabil.* **34**, 1613-1622 (2013).
22. T. T. Y. Wong, C. S. H. Ho, Component processes in arithmetic word-problem solving and their correlates. *J. Educ. Psychol.* **109**, 520-531 (2017).
23. G. Borst, N. Poirel, A. Pineau, M. Cassotti, O. Houdé, Inhibitory control in number-conservation and class-inclusion tasks: A neo-Piagetian inter-task priming study. *Cogn. Dev.* **27**, 283-298 (2012).
24. N. Gogtay, J. N. Giedd, L. Lusk, K. M. Hayashi, D. Greenstein, A. C. Vaituzis, T. F. Nugent, D. H. Herman, L. S. Clasen, A. W. Toga, J. L. Rapoport, P. M. Thompson, Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8174-8179 (2004).
25. N. C. Jordan, S. C. Levine, J. Huttenlocher, Calculation abilities in young children with different patterns of cognitive functioning. *J. Learn. Disabil.* **28**, 53-64 (1995).
26. L. S. Fuchs, D. L. Compton, D. Fuchs, K. Paulsen, J. D. Bryant, C. L. Hamlett, The prevention, identification, and cognitive determinants of math difficulty. *J. Educ. Psychol.* **97**, 493-513 (2005).
27. M. Cassotti, S. Moutier, How to explain receptivity to conjunction-fallacy inhibition training: Evidence from the Iowa gambling task. *Brain Cogn.* **72**, 378-384 (2010).
28. O. Houdé, L. Zago, E. Mellet, S. Moutier, A. Pineau, B. Mazoyer, N. Tzourio-Mazoyer, Shifting from the perceptual brain to the logical brain: The neural impact of cognitive inhibition training. *J. Cogn. Neurosci.* **12**, 721-728 (2000).
29. F. Faul, E. Erdfelder, A. Buchner, A.-G. Lang, Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav. Res. Methods* **41**, 1149-1160 (2009).