Acta Psychologica 129 (2008) 315–324

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

A generalization of the representational change theory from insight to non-insight problems: The case of arithmetic word problems

Catherine Thevenot [a,*], Jane Oakhill [b]

[a] FAPSE, University of Geneva, 40, bd du Pont D'Arve, CH-1205 Geneva, Switzerland
[b] Department of Psychology, School of Life Sciences, University of Sussex, Falmer BN1 9QH, UK

ARTICLE INFO

Article history:
Received 15 May 2008
Received in revised form 21 July 2008
Accepted 19 August 2008
Available online 2 October 2008

PsycINFO classification:
2300 Human Experimental Psychology
2340 Cognitive processes
2343 Learning and Memory

Keywords:
Word problems
Verbal problems
Strategies
Mathematical cognition

ABSTRACT

This paper provides evidence for a possible generalization of Knoblich and colleagues' representational change theory [Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1534–1555; Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem. *Memory and Cognition, 29*, 1000–1009] outside its original scope of application. While this theory has been proposed to explain insight problem solving, we demonstrate here that its main concepts, namely, constraint relaxation and chunk decomposition, are applicable to incremental problem solving. In a first experiment, we confirm, as already shown by problem solving and reasoning researchers, that individuals avoid the construction of alternative representations of the problems when possible. In the second and third experiments, we show that alternative representations of arithmetic problems are easier to construct and maintain when they violate constraints of narrow rather than wide scope. The specificity of insight problem solving is discussed in the light of these new findings.

1. Introduction

Incremental problems such I as the Tower of Hanoï, also called analytic problems, require the solver to take a series of steps in order to reach the solution. Solving the problem might take time but the solver usually has a good idea of how to reach it early in the problem-solving process. In contrast, insight problems such as the nine-dot problem are not solved step by step but rather suddenly with 'a flash of illuminance' (Metcalfe & Wiebe, 1987). In order to differentiate insight from incremental problems, three main approaches have been adopted in the literature. The first approach focuses on the processes underlying problem solving. Insight and incremental problems would rely differently on the two systems of reasoning described in the dual-process approach (Evans, 2003, 2008; Stanovitch & West, 2000). Insight problem solving would show a greater involvement of system 1 processes, which are rapid, parallel and automatic. On the contrary, incremental problem solving would depend more heavily on system 2 processes, which permit hypothetical thinking and operate slowly and sequentially (but see Kaplan and Simon (1990) for a different point of view). The second approach defines insight problems in terms of phenomenology. For example, Metcalfe and Wiebe (1987) characterize insight problems as those showing an absence of incremental feeling of warmth ratings before solution. The third approach emphasizes changes in conceptual knowledge necessary for insightful solutions to be found (i.e., changes in approach (Weisberg, 1995) or, as we will see later in detail, changes in problem representation (Knoblich, Ohlsson, & Raney, 2001, for example)).

As mentioned above, contrary to incremental problems, which are solved in a steady approach toward the goal, insight problems are often solved suddenly. This unexpected success usually occurs after a period of impasse. Two main theories account for the impasse phenomenon. The first one is 'the progress monitoring theory' proposed by MacGregor, Ormerod, and Chronicle (2001), which suggests that problem solving proceeds with the problem solver seeking to minimize the gap between the current state of the problem and the goal or sub-goal state. The impasse occurs when the problem solver finds that this hill-climbing method does not give rise to the solution, and it is only at this point that alternative approaches are considered. The second theory that accounts for the impasse phenomenon is the 'representational change theory' put forward by Knoblich and colleagues (Knoblich, Ohlsson, Haider, & Rhenius, 1999; Knoblich et al., 2001). According to the authors, the impasse occurs because the solver constructs an initial

* Corresponding author. Tel.: +41 22 379 92 52.
E-mail address: catherine.thevenot@pse.unige.ch (C. Thevenot).

0001-6918/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2008.08.008

representation of the problem that has a low probability of success. It is only by revising this representation that the solver will overcome the impasse. The probability of such a successful revision depends on two processes, namely, constraint relaxation and chunk decomposition. Within the initial representation, some constraints (i.e., what does and does not count as a solution) are inappropriate and need to be relaxed. Certain constraints are more likely to be relaxed than others. Indeed, according to the authors, the probability of relaxation of a constraint is inversely proportional to its scope. In other words, it depends on how much a problem representation is affected if the constraint is relaxed: constraints of narrow scope have a higher probability of being relaxed than constraints of wide scope because the resulting revisions in the problem representation are more circumscribed. For example, in a matchstick arithmetic problem the goal is to move a single stick in such a way that an initial false statement is transformed into a true arithmetic statement. Moving from the false equation VI = VII + I to the true one VII = VI + I requires the solver to relax a narrower constraint than does the problem III = III + III, which requires the solution III = III = III. Indeed in the latter example, the whole equation has to be revised whereas in the first example, the revision concerns only the values. The construction of the new representation supposes a deconstruction of the initial representation by a chunk decomposition process. Chunk decomposition corresponds to those objects that can be broken down into further objects. Loose chunks such as VII, which decompose into additional chunks (V and II, for example), are more likely to be decomposed than tight chunks, which decompose into entities that are not themselves chunks (*X* for example). To sum up, in Knoblich's theory, a problem can be re-represented by relaxing the unnecessary constraints that have been placed on the problem and/or decomposing chunked objects in the representation.

As pointed out by Jones (2003), the 'progress-monitoring theory' and 'the representational change theory' have different scopes of application. The first one cannot be applied to problems that can be solved in only one move because it relies on the constant monitoring of progress in the problem. It is indeed difficult to gain support for this theory for the problems that do not give the solver the sense that they are getting closer to the solution. On the contrary, the main support for the representational change theory is from one-step problems. However, despite different emphases and scope, these approaches recognize that insight problem solving involves some kind of restructuring of the initial representation. This representational change can be viewed as a specific characteristic of insight problem solving (Weisberg, 1995). However, we have shown recently that restructuring of the initial mental representation can be observed under certain circumstances when individuals have to solve arithmetic word problems (Thevenot & Oakhill, 2005, 2006), which are considered to be incremental problems. This fact strengthens Wieth and Burns' conclusion (2006) that it is possible that the problem-solving processes of incremental and insight problems are more similar than previously considered. We propose here to put this assumption to the test and attempt to generalize Knoblich and colleagues's theory beyond its original scope of application. As stated earlier, theories of insight are more or less applicable depending on the characteristics of the problems (Jones, 2003). As we will see below, arithmetic problems are particularly suitable in order to manipulate the two main mechanisms for representational change proposed by Knoblich et al. (1999), namely, constraint relaxation and chunk decomposition.

In verbal arithmetic, multiple-step problems can be solved by different strategies; all of which may lead to the solution, but not all in an optimum way. Let us consider the problem 'how many marbles does Tom have less than John and Paul together? John has 41 marbles, Tom has 31 marbles and Paul has 43 marbles'. The strategy that is most likely to be used by individuals is the one based on the sub-goal explicitly described in the text of the problem (Thevenot, Barrouillet, & Fayol, 2004; Thevenot & Oakhill, 2005, 2006). Most adults will reach the sub-goal 'John + Paul' and then subtract the quantity of Tom's marbles from the result obtained: $41 + 43 = 84 - 31 = 53$. However, another strategy would be to subtract the quantity of Tom's marbles from the quantity of John's marbles and then, to add the quantity of Paul's marbles to the result obtained: $41 - 31 = 10 + 43 = 53$. In order to solve the problem mentally, this second strategy is obviously more economical than the first one. However, it is not implemented spontaneously by individuals who base their strategy on their initial mental representation of the problem. Such individuals would not engage in the construction of an alternative representation because it is an additional and rather demanding activity. The fact that individuals do not systematically construct alternative representations when possible echoes the results obtained both in reasoning and in problem-solving studies. Indeed, mental model theorists in the domain of syllogistic and conditional reasoning showed that alternative mental models are rarely constructed when the initial model provides a solution (Evans, Handley, Harper, & Johnson-Laird, 1999; Newstead, Thompson, & Handley, 2002; Polk & Newell, 1995). The same results are obtained by problem-solving researchers who show that alternative representations are not constructed when a solution path from the initial representation is efficient (Knoblich et al., 2001; Luchins & Luchins, 1959).

Nevertheless, we showed that when the solution is extremely difficult to reach from the initial representation, an alternative one that leads to the optimum strategy can be constructed by the solvers (Thevenot & Oakhill, 2005). More concretely, by using three- instead of two-digit numbers in the text of arithmetic problems, we pushed our participants to implement a less demanding strategy than the one induced by the text of the problem (see the example above). Following Knoblich and colleagues, the initial representation of the problem had a low probability of success. Note that, in fact, the initial representation constructed by individuals in order to solve insight problems such as the matchstick arithmetic, the nine-dot or the candle problem has no probability of success at all. A difference in multiple-step arithmetic problem solving is that the initial representation can lead to the problem solution but with a lower probability of success than the alternative representation.

Following Knoblich's and colleagues, what has to be determined is which factor(s) constrained the solvers' initial representation. In the case of arithmetic word problems, it is the semantic structure of the problem that shapes up the mental representation from which the solution plan will be derived (Nescher, Greeno, & Riley, 1982; Riley, Greeno, & Heller, 1983). The problem described above: 'how many marbles does Tom have less than John and Paul together?' is a two-step problem that describes a 'combine' and a 'compare' situation. As described by Riley et al. (1983), 'combine' problems involve a static relationship between sets that are to be considered in combination (i.e., John + Paul in our example). 'Compare' problems also describe two quantities that do not change, but this time the solver is asked to determine the difference between them (i.e., Sum of John's and Tom's marbles – Tom in our example). It is very clear here that the initial mental representation of the problem constructed by the solver (i.e., '(John + Paul) – Tom') is constrained by the semantic relations between the different sets described in the text. In order to re-represent the problem and adopt a different solving strategy, the solver has to relax these constraints, or in other words disregard the semantic structure(s) of the problem and re-organize the relationships between the different sets. As the strategy resulting from the mental representation induced by the mathematical problem statement can be represented in an equation, the scope of the constraint, or in other words the distance between the two representations (i.e., before and after the relaxation of constraints), can be easily and precisely assessed. In order to change the initial representation '(John + Paul) – Tom' into the alternative one '(John – Tom) + Paul', the chunk '(John + Paul)' has to be broken down into John and Paul, then the resulting representation 'John + Paul – Tom' has to be re-organized into 'John – Tom + Paul' and finally the new chunk '(John – Tom)' has to be created. In this example, three steps separate the new and the initial representation, and the chunk decomposes into two independent and meaningful objects. As described in the introduction to the second experiment in this paper, we will focus on word problems that differ in the number of steps they require for the construction of the new representation (i.e., in the scope of the constraint), while the tightness of the chunk to be decomposed will remain constant.

In order to investigate the strategies that derived from the mental representations constructed by individuals to solve arithmetic problems, we have developed a paradigm that does not rely on verbal reports. Even if verbal reports are usually the method preferred by researchers to gather information about an individual's strategies (Coquin-Viennot & Moreau, 2003; Gonzales & Espinel, 2002; Moreau & Coquin-Viennot, 2003; Pape & Wang, 2003, for example), doubts have been expressed by cognitive psychologists concerning self-reports in general (Crutcher, 1994; Ericsson and Simon, 1980; Payne, 1994; Wilson, 1994) and in the domain of arithmetic in particular (Kirk & Ashcraft, 2001). Kirk and Ashcraft (2001) point out that verbal reports are not always accurate, can alter the mental processing that normally occurs and, finally, can draw the attention of participants to the experimenter's hypotheses. In order to avoid these biases, other researchers prefer to use eye-movement registration. This technique aims at understanding how and from which specific textual clues the mental representation of the problem is constructed. The text elements that are fixated longer are assumed to be more deeply processed (Just & Carpenter, 1980; Rayner & McConkie, 1976) and, consequently, central to the construction of the representation and the production of the solution (De Corte, Verschaffel, & Pauwels, 1990; Hegarty, Mayer, & Green, 1992; Hegarty, Mayer, & Monk, 1995; Verschaffel, De Corte, & Pauwels, 1992). However, because of the frequent regressions to previous lines and the numerous short fixations on certain elements, eye-movement registrations make it difficult to know exactly which calculations are performed for the resolution. More recently, Green, Lemaire, and Dufau (2007), who studied mental calculation strategies on complex addition using eye-movement registration, called attention to the fact that even if a large number of studies have employed this technique in different cognitive domains, there is still a lack of standardization in methods of analysis in the problem-solving literature.

In order to get around these potential methodological biases, we conceived a paradigm that takes advantage of the fact that algorithmic computation (i.e., non-retrieval) degrades the memory traces of the operands used in the calculation (Thevenot, Barrouillet, & Fayol, 2001; Thevenot, Fanget, & Fayol, 2007). By presenting the text sequentially and interrupting it for a recognition test of the different numbers previously presented, we can determine whether these numbers have already been used in calculation or are currently maintained in working memory for a forthcoming computation. In the former case, the operand memory traces are degraded and thus less accessible than in the latter case, where the memory traces are kept active in short-term memory. However, problem solvers can do calculations during reading only if they are told about, or can infer, the nature of the problem. In the problems we used, the question was the only clue that provided information about the structure of the problem. Indeed, the text of the problem was constructed in order to avoid any expressions such as "more than", "more less" or "together", which are known to potentially trigger a specific resolution scheme (De Corte et al., 1990, for example). Thus, calculation during reading could only be expected to take place when the question was provided before and not after the numerical information (Devidal, Fayol, & Barrouillet, 1997; Fayol, Abdi, & Gombert, 1987). In fact, when the question is presented at the end of the text, no calculation is possible while reading the numerical values. Therefore, if longer recognition times and higher rates of recognition errors for the operands are obtained when the question is before rather than after the text, we can conclude that calculations have been achieved during reading.

In the second experiment of this paper, we studied three types of problems that differ on the scope of the constraint to be relaxed in order to set up a more economic strategy than the one induced by their semantic structure. As explained before, these more economic strategies are not implemented spontaneously by individuals (Thevenot & Oakhill, 2005, for example), so we explicitly asked our participants to find an alternative strategy. Once the strategy found, we were interested in knowing whether or not the new strategy would be maintained. We hypothesized that the narrower the scope of the constraint to be relaxed, the more probable the maintenance of the new strategy. In a third experiment, we not only taught the participants the new strategy, but we also strongly encouraged them to use it to solve the problems. We hypothesized that the narrower the scope of the constraint to be relaxed, the quicker the automatic use of the new strategy. Through these two manipulations, we expected the participants to discover and then use a strategy that they do not spontaneously implement. However, the observation and interpretation of a new behaviour make sense only in the light of the initial behavior. Therefore, in a first experiment, we expected to replicate and extend our previous results to a type of problem that we had not studied before. The strategies (i.e., the order of completion of calculations) used by the participants should be based on their initial mental representation, which is constrained by the semantic structure of the problem.

2. Experiment 1

The three problems studied in this experiment were all of the form 'John has X marbles, Tom has Y marbles and Paul has Z marbles' but differed in their associated question. P1 problems were associated with the question "how many marbles do John and Tom together have more than Paul?", P2 problems were associated with the question "how many marbles does John have more than Tom and Paul together?" and finally, P3 problems were associated with the question "how many marbles does Tom have less than John and Paul together?". Whereas P1 and P2 problems were studied in our previous research (Thevenot & Oakhill, 2006, for example), P3 problems had never been studied before. This last problem was introduced here in order to manipulate the scope of the constraints that needed to be relaxed in order to implement an alternative strategy. As the different quantities always followed the same order (i.e., John then Tom and then Paul), only P1 problems explicitly specified a sub-goal that could be reached as soon as the quantity of Tom's marbles was presented to the participant. On the contrary, P2 and P3 problems would require restructurings to be solved during reading. These restructurings, which are broader in the case of P3 than P2 problems, constitute the focus of the subsequent experiments, and will be described in detail later.

We expect to replicate the results of our previous research and extend them to a new type of problem (i.e., P3): when the question is before the text, calculations during reading (i.e., a first calculation after the presentation of Tom's quantity of marbles) should occur only for the P1 problem, which is the only problem that specifies a sub-goal implying the quantity of John and Tom's marbles. On the contrary, participants should wait until the end of the problem (i.e., the presentation of Paul's quantity of marbles) before engaging in calculations when confronted with P2 and P3 problems, which explicitly describe a sub-goal involving the quantity of Paul's marbles. Therefore, whereas poorer performance of recognition of the operands (i.e., lower rates of correct recognition and/or longer recognition times) should be observed when the question is before rather than after the text of P1 problems, no such difference should be obtained for the P2 and P3 problems.

2.1. Method

2.1.1. Participants

Thirty-five participants took part in this experiment. They were all undergraduate volunteers at Sussex University, and they received £5 for participating.

2.1.2. Material and procedure

Each experimental trial consisted of a text, a recognition test, and a question. All the texts we used comprised of three segments, which provided numerical information. The text was always of the form "John has *X marbles,* Tom has *Y marbles,* and Paul has *Z* marbles" and was associated either with the question "how many marbles do John and Tom together have more than Paul?" (P1 problems), "how many marbles does John have more than Tom and Paul together?" (P2 problems) or "how many marbles does Tom have less than John and Paul together?" (P3 problems). The question was presented either before or after the text.

The numerical values *X*, *Y*, and *Z* were two-digit numbers. In order to optimize the probability of an algorithmic solution (required for the degradation of the memory traces), the values of operands *X* and *Y* were limited to a range from 12 to 49, their sum always required a carry, and neither of the operations (addition and subtraction) had an answer ending with a 0. The quantity of Paul's marbles (*Z*) was chosen to ensure the coherence of the problem. For the recognition test, a numeral was displayed on screen after the second (Tom's) segment. In half of the cases, this numeral corresponded to one of the two numbers presented previously in the problem (either *X* or *Y*). In the other half, this numeral was a filler constructed by adding or subtracting 1 or 2 to either *X* or *Y*, and required a 'no' response.

The three types of problems (P1, P2 and P3) and the four types of numbers involved in the recognition task (X, Y, $X \pm 1$ or 2, $Y \pm 1$ or 2) resulted in 12 problems that were presented to each participant twice with different numerical values (24 problems). Each of these 24 problems was presented with the question either before or after the text, thus resulting in 48 experimental trials.

We used a segmented self-presentation procedure in which information is displayed on screen segment by segment. For example, the successive segments took the following form: "John has 39 marbles/Tom has 17 marbles/39/and Paul has 16 marbles/how many marbles does John have more than Tom and Paul together ?" for a P2 problem with the question after the text and a numeral corresponding to *X* for the recognition test. When the participants had read a segment, they replaced it with the next segment by pressing the space bar on the keyboard. They were asked to read each problem carefully, to solve it in their head, and to answer the question out loud. Moreover, they were instructed that a target (a numeral) would appear on screen after the second numerical segment and were asked to decide whether this number had been previously presented or not. They gave their response by pressing a "yes" or "no" key. The nature of the responses and reaction times were registered from the onset of the target to the response. No feedback about the accuracy of the answers was given to the participants.

The experiment was controlled by a version of the TSCOP software (Norris, 1984) and carried out using a PC-compatible computer fitted with an Advantech PCLabCard, which provided millisecond timing of responses made via buttons attached to the card.

Table 1
Proportions of correct answers in the arithmetic task (and standard deviations) as a function of the Position of the Question and the type of problem in Experiment 1

Question position	Type of problem		
	P1	P2	P3
Before	.91	.80	.63
	(.14)	(.17)	(.24)
After	.60	.61	.51
	(.22)	(.21)	(.29)

2.2. Results

2.2.1. Proportions of correct answers in the arithmetic task

A 2 (Position of the Question: before vs. after) × 3 (type of problem: P1, P2, P3) ANOVA with the two factors as repeated measures was performed on the proportions of errors in the resolution task (Table 1).

The classical facilitatory effect of the Position of the Question before the text (Fayol et al. 1987, for example) was significant: the proportion of correct answers to the arithmetic task was higher when the question was before (.78) rather than after the text (.57), $F(1,34) = 57.56$, MSE = .040, $p < .001$. There was also a significant effect of the type of problem, $F(2,68) = 17.55$, MSE = .040, $p < .001$: P1 problems were easier (.75) than P2 problems (.70), $F(1, 34) = 5.01$, MSE = .018, $p = .03$ and P2 problems were easier than P3 problems (.57), $F(1,34) = 15.44$, MSE = .040, $p < .001$.

The interaction between the Position of the Question and the type of problem was significant, $F(2,68) = 9.87$, MSE = .018, $p < .001$. The facilitatory effect of the question before was higher for P1 than for P2 problems ($F(1,34) = 7.35$, MSE = .018, $p = .01$) and higher for P1 than for P3 problems ($F(1,34) = 17.23$, MSE = .020, $p = .001$). However, the facilitatory effect was the same for P2 and P3 problems, $F(1,34) = 3.16$, MSE = .015, $p > .05$.

2.2.2. Proportions of correct recognitions of the operands

An analogous analysis was performed on the proportions of correct answers in the recognition task of the operands (Table 2). Of the 840 (35 participants × 24 trials) relevant trials, 251 corresponded to incorrect responses to the arithmetic task and were discarded. One participant, who did not give a single correct answer for a certain category of problem, was eliminated from the analysis.

The proportions of correct recognitions were higher when the question was after (.99) rather than before (.95) the text of the problem, $F(1,33) = 8.63$, MSE = .01, $p = .006$. Moreover, the main effect of the type of problem was significant, $F(2,66) = 5.19$, MSE = .006, $p = .008$: the proportions of correct recognitions were higher for P2 and P3 problems (.98 for both problems) than for P1 problems (.94), $F(1,33) = 6.29$, MSE = .011, $p = .02$.

The interaction between the Position of the Question and the type of problem was marginally significant, $F(2,66) = 3.01$, MSE = .009, $p = .06$. As expected, the effect of the Position of the Question was significant only for P1 problems ($F(1,33) = 7.87$, MSE = .017,

Table 2
Proportions of correct recognitions of the operands (and standard deviations) as a function of the Position of the Question and the type of problem in Experiment 1

Question position	Type of problem		
	P1	P2	P3
Before	.90	.97	.97
	(.16)	(.08)	(.10)
After	.98	.99	.98
	(.06)	(.04)	(.06)

Table 3
Mean recognition times of the operands in ms (and standard deviations) as a function of the Position of the Question and the type of problem in Experiment 1

Question position	Type of problem		
	P1	P2	P3
Before	1193 (491)	988 (368)	920 (205)
After	950 (325)	955 (259)	946 (287)

$p = .008$) and not for P2 ($F(1,33) = 1.84$, MSE = .004, $p = .18$) and P3 problems ($F < 1$).

2.2.3. Recognition times of the operands

On the 589 trials included in the previous analysis, 17 in which participants failed to recognize the target were discarded, and the reaction times from the remaining 572 trials were analysed using an ANOVA with the same design as above (Table 3).

The recognition times were higher when the question was before (1034 ms) rather than after (950 ms) the text of the problem, $F(1,33) = 5.43$, MSE = 65,374, $p = .03$. The effect of the type of problem was significant, $F(2,66) = 7.65$, MSE = 45,630, $p < .001$: the recognition times were higher for P1 (1072 ms) than for P2 (971 ms) and P3 problems (933 ms), $F(1,33) = 11.88$, MSE = 54,473, $p < .001$.

More interestingly, the interaction between the Position of the Question and the type of problem was significant and followed the same pattern as the proportions of correct recognition of the operands, $F(2,66) = 5.50$, MSE = 61,726, $p = .006$: the effect of the Position of the Question was significant only for P1 problems, $F(1,33) = 8.70$, MSE = 115,406, $p = .006$ and not for P2 and P3 problems, both Fs = 1.

2.3. Discussion

The results of this experiment replicate the results of our previous research (e.g. Thevenot & Oakhill, 2005) and extend them to a new type of problem (i.e., P3): the mental representation constructed by individuals to solve a word problem reflects its semantic structure. Therefore, the sub-goal reached by individuals is the sub-goal that is explicitly described by the text of the two-step problem. Indeed, poorer recognition of the operands when the question was before rather than after the text occurred only when participants solved a problem describing a sub-goal that could be reached after the presentation of the second numeric segment (i.e., P1: how many marbles do John and Tom together have more than Paul?). No such difference in the recognition performance of the operands was observed for P2 and P3 problems, which describe a sub-goal that involves the last protagonist's quantity of marbles (i.e., Paul's). As explained before, a difference in the performance of recognition as a function of the position of the question allows us to infer that calculations are performed during reading. It therefore appears that no alternative mental representation of the problem was constructed by individuals for P2 and P3 problems and that they waited until the very end of the problem before engaging in calculations, even when the question was presented before the problem.

The results of this first experiment show, as well, that performing the calculations during reading is a good strategy. Indeed, the facilitatory effect of the Position of the Question before the text was more pronounced for P1 than for P2 or P3 problems. This result is not surprising: when calculations are performed during reading, only two quantities have to be maintained in working memory (i.e., the sub-result and the quantity of Paul's marbles). On the contrary, if no calculation is performed during reading, three quantities have to be maintained in working memory (i.e., the quantities of John, Tom and Paul's marbles). The cognitive resources spared can therefore be attributed to the achievement of calculations, hence the improvement in performance.

3. Experiment 2

We have just confirmed that, in order to solve an arithmetic word problem, adults derive their solution plan from the representation that is induced by the wording of the text problem, even when a better strategy could be implemented. Indeed, if we consider the characteristics of our paradigm, the strategy that consists in performing successive operations as soon as the numbers are presented is less cognitively demanding than a strategy that consists in waiting until the end of the problem before engaging in calculations. When this last strategy is implemented, individuals have to maintain three quantities and their respective relations in working memory until the end of the problem. On the contrary, if a first operation is achieved as soon as the quantity of Tom's marbles is known, the representation can be simplified during reading: after the first calculation, the respective quantities of John and Tom's marbles are of no further use for the resolution, and only the result of the operation has to be maintained in working memory until the end of the problem. As explained before, setting up this second strategy requires the relaxation of the constraint induced by the semantic structure of the problem.

In this second experiment, the same problems that were used in the previous experiment were again studied. However, prior to the arithmetic task, we explicitly asked the participants to find an alternative strategy to solve P2 (i.e., how many marbles does John have more than Tom and Paul together?) and P3 problems (i.e., how many marbles does Tom have less than John and Paul together?). More precisely, the participants were told that the alternative strategy must permit them to reach a sub-goal involving John and Tom's marbles. As described in the Introduction of this paper, in order to change the initial representation of a P3 problem '(John + Paul) – Tom' into the alternative one '(John – Tom) + Paul', the chunk (John + Paul), which results from the combine structure of the problem, has to be broken down in order to consider John and Paul independently. The decomposition of the chunk leads to the mental representation of the problem as follows: John + Paul – Tom (i.e., breaking the chunk (John + Tom) corresponds, in mathematical terms, to the development of the equation). Then, a reorganization into John – Tom + Paul is necessary and, finally the new chunk (John – Tom) has to be created. Three steps (i.e., chunk decomposition leading to the development of the equation, reorganization and rechunking) separate the new and the initial representation. Here, the chunk decomposes into two independent and meaningful objects. As far as the P2 problems are concerned, the execution of the calculation during reading requires a change of the initial representation 'John – (Tom + Paul)' into the alternative one '(John – Tom) – Paul'. For this transformation, the chunk '(Tom + Paul)' has to be decomposed, which leads to represent the problem as follows: John – Tom – Paul. No reorganization is needed but the new chunk (John – Tom) has to be created. Whereas the tightness of the chunk to be decomposed is exactly the same for P2 and P3, the scope of the constraint is narrower for P2, which require only two steps for the representational change, than for P3.

As a consequence and following Knoblich's theory, our first hypothesis is that it will take longer to find an alternative strategy for P3 than for P2 problems because the scope of the constraints is more circumscribed in the case of P2 than of P3 problems and the constraints is therefore more difficult to relax for P3 than for P2 problems. Once new representations of P2 and P3 problems are constructed by participants, the question that arises is whether or not these representations will continue to guide individual's strategies during the resolution task. The same paradigm as in the previous experiment was used in order to study the potential differences in the strategies used by our participants in this second experiment. If, once discovered, an optimum strategy is maintained, the calculation will be achieved during reading when the question is before the text for all problem types, and not only for P1 problems as in the previous experiment. Because algorithmic computation degrades the memory traces of the operands used in the calculation, the operands should be better recognized (higher rate of correct recognition and/or shorter recognition times) when the question is positioned at the end rather than at the beginning of all three problems. Moreover, we measured the self-presentation times of the question when it was presented before the text. Indeed, even if in the second phase of this experiment calculations are achieved during reading for all three types of problems, this strategy change might result from a more demanding mental process in the case of P3 than of P2 problems. The shifting from the initial representation to the alternative one necessarily occurs as soon as the question is presented to participants. If it is still more difficult for participants to construct this alternative representation for P3 than for P2 problems, then the question will be self-presented longer for the former than for the latter.

3.1. Method

3.1.1. Participants

Twenty-nine participants took part in this experiment. They were all undergraduate volunteers at Sussex University, and they received £7 for participating.

3.1.2. Material and procedure

3.1.2.1. The discovery of an alternative strategy task. This task was a pen and paper one. Participants were presented with both P2 and P3 problems. In half of the cases P2 was presented first and in the other half P3 was presented first. Participants were asked to write down the calculations they would perform to solve the problem they were presented with. Just after this first step, they were asked to find another way to solve it. More precisely, they were asked to write down how they would solve the problem if they were to first find a sub-result that did not involve the number of marbles that Paul has. The time required by the participant to find this alternative strategy was measured by the experimenter with a stopwatch. The watch was stopped as soon as the right answer was written down by the participant.

3.1.2.2. The arithmetic word problem-solving task. The materials and procedure were the same as in Experiment 1 but, in addition to the recognition times, the self-presentation times of the question before the text of the problem were recorded by the computer program.

3.2. Results

3.2.1. Resolution times of the problems with the alternative strategy

In the first phase of the task, all participants solved the problems by reaching the sub-goal that was explicitly described by the text of the problem. In the second phase all the participants were able to discover the alternative strategy but, in accordance with our hypothesis, it took longer for the participants to do so for P3 (134 s) than for P2 problems (83 s), $t(28) = 2.63$, $p = .01$.

3.2.2. Reading times of the questions before the text

The reading times of the questions before the text varied as a function of the type of problem, $F(2,56) = 8.45$, MSE = 19,01,450, $p < .001$. It took longer for participants to read the question when placed before P3 problems (7910 ms) than before P1 problems (6715 ms) ($F(1,28) = 10.30$, MSE = 20,10,353, $p = .003$) or P2 problems (6543 ms) ($F(1,28) = 15.74$, MSE = 17,21,679, $p < .001$). However the reading times of the questions did not differ significantly for P1 and P2 problems, $F < 1$.

Table 4
Proportions of correct answers in the arithmetic task (and standard deviations) as a function of the Position of the Question and the type of problem in Experiment 2

Question position	Type of problem		
	P1	P2	P3
Before	.92	.81	.71
	(.10)	(.14)	(.19)
After	.59	.56	.51
	(.21)	(.18)	(.20)

3.2.3. Proportions of correct answers in the arithmetic task

A 2 (Position of the Question: before vs. after) × 3 (type of problem: P1, P2, P3) ANOVA with the two factors as repeated measures was performed on the proportions of errors in the resolution task (Table 4).

The classical facilitatory effect of the Position of the Question before the text was significant: the proportion of correct answers to the arithmetic task was higher when the question was before (.81) rather than after the text (.55), $F(1,28) = 91.80$, MSE = .032, $p = .001$. There was also a significant effect of the type of problem, $F(2,56) = 16.49$, MSE = .019, $p = .001$: P1 problems were easier (.75) than P2 problems (.68), $F(1,28) = 6.65$, MSE = .021, $p = .01$, and P2 problems were easier than P3 problems (.61), $F(1,28) = 11.88$, MSE = .015, $p = .002$.

The interaction between the Position of the Question and the type of problem was significant, $F(2,56) = 4.48$, MSE = .016, $p = .02$. Whereas the facilitatory effect of the Position of the Question before was significant for all three types of problem ($F(1,28) = 88.72$, MSE = .019, $p = .001$; $F(1,28) = 44.62$, MSE = .019, $p = .001$; $F(1,28) = 21.50$, MSE = .026, $p = .001$ for P1, P2 and P3 problems, respectively), this effect was more pronounced for P1 than for P2 problems ($F(1,28) = 4.35$, MSE = .015, $p = .04$) and more pronounced for P1 than for P3 problems ($F(1,28) = 6.00$, MSE = .023, $p = .02$). However the effect was the same for P2 and P3 problems, $F(1,28) = 1.29$, MSE = .010, $p = .26$.

3.2.4. Proportions of correct recognitions of the operands

An analogous analysis was performed on the proportions of correct answers in the operand recognition task (Table 5). Of the 696 (29 participants × 24 trials) relevant trials, 223 corresponded to incorrect responses to the arithmetic task and were discarded.

The effect of the Position of the Question was significant in this analysis, $F(1,28) = 7.38$, MSE = .02, $p = .01$: the proportions of correct recognitions were higher when the question was after (.98) rather than before (.91) the text of the problem. The effect of the type of problem, $F(2,56) = 2.78$, MSE = .010, $p = .07$, and the interaction between the two variables were marginally significant,

Table 5
Proportions of correct recognitions of the operands (and standard deviations) as a function of the Position of the Question and the type of problem in Experiment 2

Question position	Type of problem		
	P1	P2	P3
Before	.86	.95	.93
	(.22)	(.10)	(.16)
After	.98	.98	.96
	(.06)	(.06)	(.09)

Table 6
Mean recognition times of the operands in ms (and standard deviations) as a function of the Position of the Question and the type of problem in Experiment 2

Question position	Type of problem		
	P1	P2	P3
Before	1394	1139	1092
	(586)	(315)	(352)
After	1056	1025	955
	(331)	(272)	(284)

$F(2,56) = 2.57$, MSE = .014, $p = .08$: whatever the problem, the proportion of correct recognitions was higher when the question was after, rather than before, the text, but this effect was more pronounced for P1 than for P2 and P3 problems.

3.2.5. Recognition times of the operands

On the 473 trials kept for the previous analysis, 23 in which participants failed to recognize the target were discarded, and the reaction times from the remaining 450 trials were analysed using an ANOVA with the same design as above (Table 6).

The recognition times were higher when the question was before (1208 ms) rather than after the text (1012 ms), $F(1,28) = 14.63$, MSE = 114,593, $p = .001$. The main effect of the type of problem was significant, $F(2,56) = 10.5$, MSE = 59,543, $p = .001$. Indeed, the recognition times of the operands were higher for P1 problems (1225 ms) than for P2 (1083 ms) and P3 problems (1023 ms), $F(1,28) = 11.36$, MSE = 11,48,765, $p = .002$.

More interestingly, the interaction between the Position of the Question and the type of problem was significant, $F(2,56) = 4,43$, MSE = 49,899, $p = .02$. In contrast to Experiment 1, the recognition times were longer when the question was before rather than after each of the three types of problems, $F(1,28) = 10.74$, MSE = 154,423, $p = .003$; $F(1,28) = 5.68$, MSE = 32,755, $p = .02$; $F(1,28) = 10.08$, MSE = 27,212, $p = .004$ for P1, P2 and P3 problems respectively. However, this effect was more pronounced for P1 than for P2 ($F(1,28) = 5.44$, MSE = 67,465, $p = .03$) and more pronounced for P1 than for P3 problems ($F(1,28) = 4.10$, MSE = 71,193, $p = .05$). The effect of the Position of the Question was the same for P2 and P3 problems, $F = 1$.

3.3. Discussion

According to the hypothesis based on Knoblich's representational change theory, we showed in this experiment that alternative representations of arithmetic problems are easier to construct when they violate narrow rather than wide constraints. In effect, it took longer for our participants to find a strategy that specified a sub-goal that involved John and Tom's marbles for P3 (how many marbles does Tom have less than John and Paul together?) rather than for P2 problems (how many marbles does John have more than Tom and Paul together?).

The performance on the operand recognition task suggests that the discovery of the new strategy had an effect on the subsequent way individuals solved the problems. Indeed, operand recognition performance was lower when the question was at the beginning rather than at the end of the three problems. We can infer from this fact that calculations were achieved during reading for all types of problems and consequently that the better strategy discovered in the first phase of the experiment was maintained by our participants. It is not speculative to state that the alternative strategy is a better strategy than the one induced by the wording of the problem. Indeed, in this second experiment the facilitatory effect on performance of the question before the text was higher for P2 and P3 problems than in the first experiment (+ .19 and + .12 of correct answers for P2 and P3 problems in the first experiment and + .25 and + .20 in the second).

Despite the fact that differences in operand recognition as a function of question position were observed for the three problems studied, these differences were still greater for P1 than for P2 and P3 problems. This result suggests that the strategy consisting in achieving calculation during reading was less often used for the problems that require an alternative representation. A closer look at our data confirms this interpretation and reveals that 23 participants of 29 showed longer recognition times for the operands for P1 problems when the question was before the text, whereas only 16 participants showed this difference for both P2 and P3 problems[1], both $\chi^2(1) = 3.84$, $p = .05$. The non-systematic use of a more efficient strategy by some participants will be addressed in the General discussion.

Finally, the longer reading time for the questions in P3 than in P2 problems is of interest. Although the two problems are both solved during reading in the second phase of the experiment, this last fact is evidence that the strategy change still results from a more demanding mental process in the case of P3 than of P2 problems. This suggests that once a constraint with a wider scope is relaxed, it remains demanding to relax it again. This suggestion will be investigated in the next experiment.

4. Experiment 3

According to Knoblich et al.'s theory, constraints of wider scope are more difficult to relax than constraints of narrow scope. Moreover, the authors postulate that "once a problem representation has been changed, the change should persist and [...] differences in initial difficulty due to the need to relax constraints should disappear" (1999, p. 1535). However, the results of the previous experiment suggest that even when a constraint has been relaxed, the relaxation is still demanding when individuals are again presented with the same problem. If that is the case, it should take longer for wider constraints to be relaxed automatically. This hypothesis is investigated in this third experiment. We taught our participants how to relax the constraint of P2 and P3 problems and asked them explicitly to base their subsequent solution strategy on their new representation. We predicted that it would be more demanding to re-relax constraints for the wide constraint P3 problems than for the narrow constraint P2 problems, and that this difference would decrease during the course of the experiment, as automatization takes place. As in the previous experiment, the relative demand of the representational change was inferred from the reading times of the questions.

4.1. Method

4.1.1. Participants

Twenty-nine participants took part in this experiment. They were all undergraduate volunteers at Sussex University, and they received £5 for participating.

4.1.2. Material and procedure

Narrow constraint P2 problems (how many marbles does John have more than Tom and Paul together?) and wider constrained P3 problems (how many marbles does Tom have less than John and Paul together?) were studied in this experiment. As in the pre-

[1] Only participants who showed a difference of more than 80 ms were considered as showing longer recognition times when the question was before rather than after the text. The 80 ms limit was chosen because it corresponded to the average of negative differences. Because a negative difference cannot reflect any psychological phenomenon in our experiments, we postulated that a positive difference of 80 ms or less might, also, reflect only random variation.

vious experiments, the problems were presented segment by segment on a computer screen. The problems were presented in five blocks of eight problems (i.e., four P2 problems and four P3 problems). For the first block, P2 and P3 problems were presented alternately. For half of the participants, a P2 problem was presented first and for the other half, a P3 problem was presented first. The question was always presented before the problem and the participant had to give his or her answer when the quantity of Paul's marbles was on the screen (i.e., always the last segment). After the participant gave an answer, the problem appeared in its entirety on the screen, followed by the sentence "how did you solve the problem?" The participants had to report the calculations they had made in order to solve the problem. They did not have any instructions on the way to solve the first six problems. However, for the last problems in the first block, they were told that "there is an easier way to solve the problem by achieving a first calculation as soon as the quantity of Tom's marbles is presented on the screen". They were instructed to use this new strategy for the subsequent problems.

The four experimental blocks were then presented. In each of these blocks, the eight problems were presented in a random order. Each time the participant could not remember the new strategy or reported a non-adequate strategy, they were reminded how to solve the problem during reading.

In this experiment, our participants were strongly encouraged to use a non-spontaneous strategy. As a consequence, verbal reports were not problematic and we did not need to use the operand recognition paradigm. Indeed, the biases possibly associated with verbal reports are observed only when natural and spontaneous behaviours are studied. As noted earlier, verbal reports can alter the mental processing that normally occurs. Moreover, as our participants were explicitly asked to use a specific strategy, the object of the verbal report requirement was transparent, and could not reveal any more information to the participants about the purpose of the experiment, which had been made explicit to them.

The strategies used by the participants, as well as the answers to the problems, were collected by the experimenter, and the self-presentation times of the questions were recorded by the computer program.

4.2. Results

4.2.1. Proportions of correct answers in the arithmetic task

A 2 (type of problem: P2 and P3) × 4 (block: B1, B2, B3, B4) ANOVA with the two factors as repeated measures was performed on the proportions of correct answers in the arithmetic task (Table 7).

P2 problems (.87) were not solved more successfully than P3 problems (.86), $F < 1$. However, there was an effect of the block, $F(3,84) = 4.48$, MSE = .03, $p < .006$: the proportion of correct answers was higher in the last three blocks (.88, .88 and .89 for blocks 1, 2 and 3, respectively) than in the first block (.79). The interaction between the type of problem and the block was not significant, $F < 1$.

4.2.2. Proportions of use of the new strategies

The same analysis was performed on the proportions of use of the new strategies (Table 8). Each time the participant verbally reported a strategy that did not correspond to the one previously taught, a score of 0 was given. On the contrary, when, as taught, the reported strategy involved a sub-goal implying John and Tom's marbles, a score of 1 was given.

The new strategy taught was more often used for P2 (.95) than for P3 problems (.91), $F(1,28) = 7.30$, MSE = .01, $p = .01$. Moreover, the new strategies were more often used in the last three blocks (.97, .97 and .99 for blocks 2, 3 and 4, respectively) than in the first block (.79), $F(3,84) = 15.62$, MSE = .03, $p = .001$. Finally, the interaction between the two variables was significant, $F(3,84) = 4.04$, MSE = .01, $p = .01$: whereas in the first block the use of the new strategies was higher for P2 (.84) than for P3 problems (.73), it was quite similar in the last three blocks (.96 vs. .95, .98 vs. .96 and 1 vs. .99 for P2 and P3 problems for blocks 2, 3 and 4, respectively).

4.2.3. Mean self-presentation times of the question

The same analysis as previously presented was performed on the self-presentation times of the question (Table 9).

The self-presentation times were shorter for P2 (5358 ms) than for P3 problems (6285 ms), $F(1,28) = 6.07$, MSE = 819,90,277, p = .02. They also varied as a function of the block (7591 ms, 5654 ms, 5431 ms and 4611 ms for blocks 1, 2, 3 and 4, respectively), $F(3,84) = 15.40$, MSE = 59,96,591, p < .001. More importantly there was an interaction between the type of problem and the block, F(3,84) = 3.02, MSE = 43,68,339, p = .03: planned comparisons showed that the difference in the times was significant only for the first block (6432 vs. 8750 ms for P2 and P3 problems), $F(1,28) = 5.22$, MSE = 149238E2, $p = .03$ but was not significant for the subsequent blocks (5342 and 5966 ms for block 2, $F(1,28) = 2.62$, MSE = 21,55,384, p = .12; 5122 and 5740 ms for

Table 7
Proportions of correct answers in the arithmetic task (and standard deviations) as a function of the type of problem and the block in Experiment 3

Block	Type of problem	
	P2	P3
B1	.81	.77
	(.21)	(.23)
B2	.89	.88
	(.22)	(.17)
B3	.88	.89
	(.17)	(.20)
B4	.89	.89
	(.17)	(.24)

Table 8
Proportions of use of the new strategy (and standard deviations) as a function of the type of problem and the block in Experiment 3

Block	Type of problem	
	P2	P3
B1	.84	.73
	(.23)	(.30)
B2	.96	.95
	(.13)	(.17)
B3	.98	.96
	(.06)	(.11)
B4	1	.99
	(0)	(.05)

Table 9
Mean self-presentation times of the question in ms. (and standard deviations) as a function of the type of problem and the block in Experiment 3

Block	Type of problem	
	P2	P3
B1	6433	8750
	(2927)	(6166)
B2	5342	5966
	(2536)	(3699)
B3	5122	5740
	(3369)	(3670)
B4	4538	4684
	(2850)	(2914)

block 3, $F(1,28) = 2.28$, MSE = 24,31,962, p = .14; 4538 and 4684 ms for block 4, $F < 1$).

4.3. Discussion

The results show that even when participants are explicitly instructed to use a new strategy to solve the problem, they cannot implement it systematically. Reconstructing the strategy taught is obviously demanding. Moreover, the demand is higher when the constraint to relax for the representational change is wider. Indeed, within the first block of presentation, not only were P2 problems more often solved with the new strategy than P3 problems, but also the reading times for the questions were longer for P3 than for P2 problems.

However, after a few presentations of the same problems these differences disappeared, which suggests that the use of the new strategy can be automatized quickly. The fact that the new strategy was no more demanding after the first block is attested to by the fact that in the previous experiment there was no difference in the reading times of the question between P1 and P2 problems. As the strategy for P1 problems (i.e., sub-goal 'John + Tom' for 'how many marbles do John and Tom have more than Paul') does not require a representational change, we can infer that the new strategy for P2 is implemented directly as soon as the question is read.

5. General discussion

This paper provides evidence for a possible generalization of Knoblich and colleagues' representational change theory outside its original scope of application. While this theory has been proposed to explain insight problem solving, we have shown here that its main concepts, namely, constraint relaxation and chunk decomposition, are applicable to incremental problem solving. Our results reveal that, in the same way as Knoblich predicted in the domain of insight problems, alternative representations of arithmetic problems are easier to construct when they violate constraints of narrow rather than wide scope.

In a first experiment we confirmed that individuals avoid the construction of alternative representations when possible, which echoed the findings from our previous studies (e.g., Thevenot & Oakhill, 2005) and the results obtained both in reasoning (e.g., Evans et al., 1999) and in problem-solving studies (e.g., Luchins & Luchins, 1959). Our participants based their solution plan on the mental representation directly induced by the semantic structure of the problem (J − (T + P) for 'how many marbles does John have more than Tom and Paul together', for example). In the second and third experiments we showed that a representational change is more difficult and takes longer when it requires the relaxing of wide rather than narrow scope constraints. We as well showed that once the representational change has occurred, individuals can base their subsequent solutions on this new representation. However, the change is still demanding and the extent of this demand also depends on the scope of the constraint. Finally, after a few presentations of the same problem, the representational change demand decreases and becomes negligible.

The cognitive cost of the representational change could explain effects such as error perseveration or incorrect strategy reversing. We know that in insight problem solving, participants can realise at a certain point that they are adopting an inappropriate strategy and can adopt a better one (Ash & Wiley, 2006; Kaplan & Simon, 1990; Kershaw & Ohlsson, 2004; Schooler, Ohlsson, & Brooks, 1993). However, there is evidence that individuals can temporarily lose such "insight" and can revert to their previous strategy. Moreover, the more complex the task, the more frequent these reversions (Johnson-Laird & Wason, 1970; Oakhill & Johnson-Laird, 1985). The results of this research allow us to put forward a new explanation for this effect. Individuals could revert to their previous strategy because, when a constraint is relaxed, the relaxation is not subsequently automatic. When the task is complex, its cognitive load is high and the representational change is necessarily experienced as more demanding. The cognitive load of the task cannot be reduced and so the only way to free mental resources is to revert to the initial strategy. Our results show that the cost of the construction of alternative representations decreases after several presentations of the same problem, and we can expect such a loss of insight to decrease with practice.

The fact that, even after the relaxation of constraints, the use of a newly discovered strategy is still demanding appears at odds with Knoblich et al. (1999), who predicted that after constraint relaxation, the change should persist and, more importantly, differences in initial difficulty due to the need to relax constraints should disappear. However, Kershaw and Ohlsson (2004) argue that for specific problems, for example, problems involving perceptual and process factors, this prediction does not follow. The discrepancy between the results of Knoblich and our own can also be explained by two main differences between classical studies of insight problems and this research. First, in our experiments, the representational changes always occur after an explicit request by the experimenter. On the contrary, even if the participants are sometimes trained (Kershaw & Ohlsson, 2004, for example) or given a part of the solution (Weisberg & Alba, 1981, for example), in most of the studies on insight problem, the insight 'comes naturally' to the solver's mind. In fact, one of the questions in the insight literature is whether representational change occurs because people explicitly search for a new representation or because implicit processing triggers representational changes. A possible answer to this question could be obtained by comparing individuals' performance on insight problems when they are asked or not for such an explicit search.

As mentioned before, the second difference between classical studies on insight and our own concerns the necessity of the representational change. In our experiments, even the initial mental representation can lead to the problem solution, and the representational change simply allows our participant to set up a better strategy. On the contrary, Gilhooly and Murphy (2005) argue that in all cases the 24 "presumed insight tasks" they review can only be solved by a representational change. However, it is not to say that all insight problems necessitate the insight to occur in order to be solved. Indeed, according to Siegler (2000) and Siegler and Stern (1998), some problems can be solved in either insightful or non-insightful ways: on inversion problems of the form A + B − B, the answer is always A. The non-insightful way to solve it is to use the standard procedure of adding the first two numbers and subtracting the third. According to Siegler, the insightful way involves by simply saying the first number. Therefore, from his point of view, some problems can be qualified as insight or non-insight problems depending on the strategy chosen by the solver. In fact, the author qualifies these problems as "arithmetic insight problems". But if we adopt this point of view, all problems that can be solved with a shortcut strategy are insight problems. As a consequence, a clear boundary between insight and strategy shifts in problem-solving literature would not necessarily exist. Moreover, because almost all incremental problems can be solved by different strategies, the view that insight and non-insight solutions are attained with the same cognitive mechanisms would be largely confirmed (Wieth & Burns, 2006).

Therefore, this paper demonstrates that the problem-solving processes of incremental and insight problems are more similar than they are usually considered to be. We have shown that it is possible to use theories of insight problem solving to predict individuals' behaviour in the domain of incremental problem solving. It is important to note that our conclusions do not apply only to the problems we studied here. Whenever a better strategy or a different than the one initially set up by individuals is available, the distance between the two mental representations can be assessed. As explained before, arithmetic problems are particularly suitable for evaluating this distance because the mental representations can be represented in mathematical equations. Therefore, the width of the constraint(s) to be relaxed can be precisely evaluated. Finally, we have shown that original results obtained in the domain of incremental problem solving can be of use for the interpretation of an individual's behaviour observed in insight problem solving.

Acknowledgments

This work was supported by an ESRC Project Grant (Ref. RES-000-23-0476).

We would like to thank Stellan Ohlsson for his suggestions and comments on a first version of the manuscript.

References

Ash, I. K., & Wiley, J. (2006). The nature of restructuring in insight: An individual differences approach. *Psychonomic Bulletin and Review, 13*, 66–73.

Coquin-Viennot, D., & Moreau, S. (2003). Highlighting the role of the episodic situation model in the solving of arithmetical problems. *European Journal of Psychology of Education, 3*, 267–279.

Crutcher, R. J. (1994). Telling what we know: The use of verbal report methodologies in psychological research. *Psychological Science, 5*, 241–244.

De Corte, E., Verschaffel, L., & Pauwels, A. (1990). Influence of the semantic structure of word problems on second graders' eye movements. *Journal of Educational Psychology, 82*, 359–365.

Devidal, M., Fayol, M., & Barrouillet, P. (1997). Stratégies de lecture et résolution de problèmes arithmétiques. *L'Année Psychologique, 97*, 9–31.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215–251.

Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Science, 7*, 454–459.

Evans, J. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255–278.

Evans, J. St. B. T., Handley, S. J., Harper, C. N. J., & Johnson-Laird, P. N. (1999). Reasoning about necessity and possibility: A test of the mental model theory of deduction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(6), 1495–1513.

Fayol, M., Abdi, H., & Gombert, J. E. (1987). Arithmetic problem formulation and working memory load. *Cognition and Instruction, 4*, 183–202.

Gilhooly, K. J., & Murphy, P. (2005). Differentiating insight from non-insight problems. *Thinking and Reasoning, 11*, 279–302.

Gonzales, J., & Espinel, I. (2002). Strategy choice in solving arithmetic word problems: Are there differences between students with learning disabilities, G–V poor performance and typical achievement students. *Learning Disability Quarterly, 25*, 113–122.

Green, H., Lemaire, P., & Dufau, S. (2007). Eye movement correlated of younger and older adults' strategies for complex addition. *Acta Psychologica, 125*, 257–278.

Hegarty, M., Mayer, R. E., & Green, C. E. (1992). Comprehension of arithmetic word problems: Evidence from students' eye fixations. *Journal of Educational Psychology, 84*, 76–84.

Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology, 87*(1), 18–32.

Johnson-Laird, P. N., & Wason, P. C. (1970). Insight into a logical relation. *Quarterly Journal of Experimental Psychology, 22*, 49–61.

Jones, G. (2003). Testing two theories of insight. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29*, 1017–1027.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329–354.

Kaplan, C. A., & Simon, H. A. (1990). In Search of insight. *Cognitive Psychology, 22*, 374–419.

Kershaw, T., & Ohlsson, S. (2004). Multiple causes of difficulty in insight: The case of the nine-dot problem. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*, 3–13.

Kirk, E. P., & Ashcraft, M. K. (2001). Telling stories: The perils and promise of using verbal reports to study math strategies. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 157–175.

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1534–1555.

Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem. *Memory and Cognition, 29*, 1000–1009.

Luchins, A. S., & Luchins, E. H. (1959). *Rigidity of behavior: A variational approach to the effect of Einstellung*. Eugene: University of Oregon Press.

MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (2001). Information processing and insight: A process model of performance on the nine-dot and related problems. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 176–201.

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Memory and Cognition, 15*, 238–246.

Moreau, S., & Coquin-Viennot, D. (2003). Comprehension of arithmetic word problem by fifth-grade pupils: Representation and selection of information. *British Journal of Educational Psychology, 73*, 109–121.

Nescher, P., Greeno, J. G., & Riley, M. S. (1982). The development of semantics categories for addition and subtraction. *Educational Studies in Mathematics, 13*, 373–394.

Newstead, S. E., Thompson, V. A., & Handley, S. J. (2002). Generating alternatives: A key component in human reasoning? *Memory and Cognition, 30*(1), 129–137.

Norris, D. G. (1984). A computer-based programmable tachistoscope for nonprogrammers. *Behavior Research Methods, Instruments and Computers, 16*, 25–27.

Oakhill, J. V., & Johnson-Laird, P. N. (1985). Rationality, memory and the search of counterexamples. *Cognition, 20*, 79–94.

Pape, S. J., & Wang, C. (2003). Middle school children's strategic behavior: Classification and relation to academic achievement and mathematical problem solving. *Instructional Science, 31*, 419–449.

Payne, J. W. (1994). Thinking aloud: Insights into information processing. *Psychological Science, 5*, 241–248.

Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review, 102*, 533–566.

Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? *Vision Research, 16*, 829–837.

Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem solving ability in arithmetic. In H. P. Ginsburg (Ed.), *The development of mathematical thinking*. New York: Academic Press.

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadow insight. *Journal of Experimental Psychology: General, 122*, 166–183.

Siegler, R. S. (2000). Unconscious insights. *Current Directions in Psychological Science, 9*, 79–83.

Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology: General, 127*, 377–397.

Stanovitch, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences, 23*, 645–726.

Thevenot, C., Barrouillet, P., & Fayol, M. (2001). Algorithmic solution of arithmetic problems and operands-answer associations in LTM. *Quarterly Journal of Experimental Psychology-A, 54*, 599–611.

Thevenot, C., Barrouillet, P., & Fayol, M. (2004). Mental representation and procedures in arithmetic word problems: The effect of the position of the question. *L'Année Psychologique, 104*, 683–699.

Thevenot, C., Fanget, M., & Fayol, M. (2007). Retrieval or non-retrieval strategies in mental addition? An operand-recognition paradigm. *Memory and Cognition, 35*(6), 1544–1552.

Thevenot, C., & Oakhill, J. (2005). The strategic use of alternative representation in arithmetic word problem solving. *Quarterly Journal of Experimental Psychology-A, 58*, 1311–1323.

Thevenot, C., & Oakhill, J. (2006). Representations and strategies for solving dynamic and static arithmetic word problems: The role of working memory capacities. *European Journal of Cognitive Psychology, 18*, 756–775.

Verschaffel, L., De Corte, E., & Pauwels, A. (1992). Solving compare problems: An eye movement test of Lewis and Mayer's consistency hypothesis. *Journal of Educational Psychology, 84*, 85–94.

Weisberg, R. W. (1995). Prolegomena to theories of insight in problem solving: A taxonomy of problems. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight*. Cambridge, (MA): MIT Press.

Weisberg, R. W., & Alba, J. W. (1981). An examination of the alleged role of "fixation" in the solution of several insight problems. *Journal of Experimental Psychology:General, 110*, 169–1992.

Wieth, M., & Burns, B. D. (2006). Incentives improve performance on both incremental and insight problem solving. *Quarterly Journal of Experimental Psychology, 59*, 1378–1394.

Wilson, T. D. (1994). The proper protocol: Validity and completeness of verbal reports. *Psychological Science, 5*, 249–252.