Journal of Experimental Psychology:
Learning, Memory, and Cognition
1999, Vol. 25, No. 6, 1534–1555

0278-7393/99/$3.00

Constraint Relaxation and Chunk Decomposition in Insight Problem Solving

Günther Knoblich
Max-Planck-Institut für psychologische Forschung

Stellan Ohlsson
University of Illinois at Chicago

Hilde Haider and Detlef Rhenius
Universität Hamburg

Insight problem solving is characterized by impasses, states of mind in which the thinker does not know what to do next. The authors hypothesized that impasses are broken by changing the problem representation, and 2 hypothetical mechanisms for representational change are described: the relaxation of constraints on the solution and the decomposition of perceptual chunks. These 2 mechanisms generate specific predictions about the relative difficulty of individual problems and about differential transfer effects. The predictions were tested in 4 experiments using matchstick arithmetic problems. The results were consistent with the predictions. Representational change is a more powerful explanation for insight than alternative hypotheses, if the hypothesized change processes are specified in detail. Overcoming impasses in insight is a special case of the general need to override the imperatives of past experience in the face of novel conditions.

Experience is both a help and a hindrance. On the one hand, human beings have no choice but to consider each new situation, task, or problem in light of past experience. There is no other resource for understanding the present and anticipating the future. On the other hand, life is complex, and there is no guarantee that tomorrow will be like yesterday. Past experience is necessarily misleading part of the time.

Problem solving often unfolds in a way that reflects the need to overcome the imperatives of past experience. The thinker begins by exploring the approaches to the problem suggested by past experience. When success does not follow, he or she enters an impasse, a state of mind that is accompanied by a subjective feeling of not knowing what to do and the cessation of overt problem-solving behavior. Continued attention to the problem sometimes leads to the appearance of a new idea, solution, or approach in consciousness. If the insight turns out to be unhelpful, the impasse continues. However, if the new idea does point the way to a solution, goal attainment is likely to be purposeful and swift, in marked contrast to the hesitation and passivity of the impasse phase.

Although the exploration–impasse–insight–execution sequence was first identified in the context of scientific discovery (Helmholz, 1896; Poincaré, 1908/1952), it can be reliably elicited by relatively simple problems. It was introduced into laboratory research on problem solving by the Gestalt psychologists (Duncker, 1945; Koffka, 1935; Köhler, 1925; Wertheimer, 1945/1959; see Ohlsson, 1984a, for a review). Wallas (1926) independently arrived at a similar formulation. After a hiatus in the 1960s and 1970s, a resurgence of interest in insight has occurred in the past decade (Bowers, Regehr, Balthazard, & Parker, 1990; Dorfman, Shames, & Kihlstrom, 1996; Fiore & Schooler, 1998; Kaplan & Simon, 1990; Keane, 1989, 1996; Lung & Dominowski, 1985; Metcalfe, 1986a, 1986b; Metcalfe & Wiebe, 1987; Ohlsson, 1984a, 1984b, 1990, 1992b; Patalano & Seiffert, 1994; Schooler, Ohlsson, & Brooks, 1993; Simonton, 1988; Smith, Ward, & Finke, 1995; Sternberg & Davidson, 1995; Yaniv & Meyer, 1987).

The insight sequence cries out for explanation. If a problem is solved eventually, then the problem solver was, by definition, competent to solve it. If so, why does he or she encounter an impasse? In contrast, if he or she initially encounters an impasse, how is the problem solved? What

Günther Knoblich, Arbeitsgruppe Kognition und Handlung, Max-Planck-Institut für psychologische Forschung, Munich, Germany; Stellan Ohlsson, Department of Psychology, University of Illinois at Chicago; Hilde Haider and Detlef Rhenius, Psychologisches Institut II, Universität Hamburg, Hamburg, Germany.

This research was supported, in part, by a grant from the German Research Foundation (Graduiertenkolleg Kognitionswissenschaft, Hamburg, Germany). Experiment 2 was conducted during a visit by Günther Knoblich to the University of Illinois at Chicago. The visit was made possible by a grant from the German Academic Exchange Service (HSPII/AUFE). This research was also supported, in part, by a grant from the Cognitive Science Program of the Office of Naval Research. The opinions expressed in this article are not necessarily those of the sponsoring agencies, and no endorsement should be inferred.

We thank Mark Keane, Janet Metcalfe, and Steven Smith for their helpful comments; Fredrik Wartenberg, Barbara Kaup, and Constanze Clauser for their helpful discussions; and Silke Wehr, Andrew Halpern, Irmgard Hagen, Günther Heilmeier, and Eva Seigerschmidt for their help in collecting the data.

Correspondence concerning this article should be addressed to Günther Knoblich, Max-Planck-Institut für psychologische Forschung, Amalienstrasse 33, 80799 Munich, Germany. Electronic mail may be sent to knoblich@mpipf-muenchen.mpg.de.

mental mechanism can explain both the occurrence of an impasse and its eventual resolution (Ohlsson, 1984b)?

In previous work, Ohlsson (1992b) proposed that past experience biases the initial representation of a problem or a situation in particular ways. The initial representation activates potentially useful knowledge elements (categories, chunks, concepts, constraints, methods, operators, procedures, rules, schemas, etc.). These knowledge elements implicitly define a space of possible solutions. If past experience is not helpful vis-à-vis the problem, that initial problem space does not contain a workable solution, and an impasse will result.

The impasse can be broken by changing the representation of the problem. A new problem representation shifts the distribution of activation over long-term memory, possibly activating dormant but relevant knowledge elements. The effect is to alter or extend the space of possibilities that are considered. If the new problem space contains a solution and if that solution is short and simple to execute, the solution will be completed quickly and unhesitantly (Ohlsson, 1984b, 1992b).

To turn this idea into a precise and testable theory, we must describe how, by which processes, problem representations are changed. The purpose of this article is to present empirical evidence for two such processes: constraint relaxation (Knoblich & Haider, 1996; Ohlsson, 1992b) and chunk decomposition (Ohlsson, 1990). In this article, we define these processes, predict relative problem difficulty and differential transfer effects in the domain of matchstick arithmetic, report four experiments that tested those predictions, and discuss alternative explanations for our results.

General Theory

Constraint Relaxation

To understand a problem is, in part, to understand what does and does not count as a solution. Knowledge of this sort consists of constraints (Mitrovich & Ohlsson, in press; Ohlsson, 1992a, 1996; Ohlsson & Rees, 1991). For example, opening a door is normally subject to the constraint that the door should not become damaged in the process.

Faced with an unfamiliar problem, the problem solver knows neither what to do nor what to avoid. If the problem reminds him or her of some problem encountered in the past, constraints relevant to that problem are likely to be activated. Those constraints may or may not be adaptive vis-à-vis the unfamiliar problem (Anderson, 1989; Isaak & Just, 1995; Keane, 1996; Ohlsson, 1992b; Richard, Poitrenaud, & Tijus, 1993). If not, then the space of options circumscribed by the set of active constraints might not contain any solution to the current problem.

Impasses caused by an overly constrained solution space can be broken by relaxing the inappropriate constraints (Isaak & Just, 1995; Knoblich & Haider, 1996; Knoblich & Wartenberg, 1998; Richard et al., 1993; Shultz & Lepper, 1996). In an emergency, it might be necessary to break through a locked door. In this type of situation, problem solving might be less a matter of searching among possibilities than of redefining what to search for. To break through a locked door in time, one should perhaps search for an ax rather than a key.

We do not claim that constraint relaxation is deliberate or voluntary. Instead, we propose that expanding the set of options to be considered by relaxing constraints is one of the mind's responses to persistent failure.

Task difficulty. At any one moment in time, several constraints are likely to be active, and they are unlikely to be relaxed all at once. It is more plausible that some constraints have a higher probability of being relaxed than others.

We hypothesize that the probability that a constraint be relaxed is inversely proportional to its scope. The scope of a constraint is determined by how much a problem representation is affected if that constraint is relaxed. If problem representations are analogous to parse trees and if a constraint applies at some node in such a tree, then scope is determined by the height of that node. A constraint has a wider scope the higher the node at which it applies. Cognitive economy dictates that constraints of narrow scope should have a higher probability of being relaxed than constraints of wide scope because the resulting revisions in the problem representation are more circumscribed. The idea that knowledge structures tend to undergo local, peripheral, or superficial changes before they undergo global, central, or fundamental changes has been proposed in cognitive psychology (Chi, 1992), developmental psychology (Vosniadou, 1994), and social psychology (Rokeach, 1970).

To use the constraint relaxation hypothesis to predict the relative difficulty of individual problems, one needs to (a) identify the constraints that apply in each problem, (b) rank order the constraints with respect to scope, and (c) identify which constraints need to be relaxed to solve the problems. Everything else being equal, a solution that violates narrow constraints should be easier to think of than a solution that violates wide constraints, and a solution that violates few constraints should be easier to think of than one that violates several constraints.

Transfer. Once a problem representation has been changed, the change should persist and so should transfer to all relevant subsequent problems. Hence, differences in initial difficulty due to the need to relax constraints should disappear. That is, all problems should become as easy as the simplest problem.

Chunk Decomposition

Familiarity with a class of objects or events leads to the creation of patterns that capture recurring constellations of features or components. This process is generally referred to as *chunking*. Expert problem-solving performance is associated with a large repertoire of problem-relevant chunks (Ericsson & Lehmann, 1996; Holding, 1985, pp. 113–120).

When faced with an unfamiliar task, the problem solver cannot know which chunks acquired in past experience are relevant for the solution. However, the application of perceptual chunks is an automatic process (Brewer, 1988; Devine, 1989; Purdue & Gurtman, 1990). If a chunk is available in a person's memory, he or she will automatically recognize instances of that chunk in the environment.

If the available chunk repertoire does not parse the problem situation in a way that is helpful vis-à-vis finding the solution, an impasse might result. This type of impasse can be broken by decomposing the inappropriate chunks into their component features and so paving the way for an alternative parse of the problem situation (Knoblich & Wartenberg, 1998; Ohlsson, 1990). Researchers have studied the formation of chunks (Anderson, 1993; Koedinger & Anderson, 1990; Rosenbloom, Laird, & Newell, 1989), but there are no models of chunk decomposition in the cognitive literature.

We do not claim that chunk decomposition is under conscious, voluntary control. Rather, we suggest that moving to a more fine-grained perceptual representation by breaking up familiar perceptual patterns (chunks) is one of the mind's responses to persistent failure.

Task difficulty. When a thinker studies a problem situation, many chunks are likely to be active. If the problem solver enters an impasse, the perceptual system cannot unpack all chunks into their components at once. Presumably, the result of such a process would be experienced subjectively as a state of kaleidoscopic confusion or stimulus overload. It is more probable that some chunks are decomposed before others.

We propose that the probability that a chunk will be decomposed is inversely proportional to the tightness of the chunk. We distinguish between tight and loose chunks on the basis of whether their components are themselves chunks, that is, meaningful perceptual patterns. A chunk with components that are themselves chunks is loose; a chunk with components that are not themselves chunks is tight.

For example, experienced chess players perceive a chessboard configuration in terms of perceptual patterns that have meaning within the game (Chase & Simon, 1973; Holding, 1985). This does not imply that chess players have lost their ability to perceive individual chess pieces. Each piece remains a meaningful unit by itself even though it appears as a component of one or more familiar patterns. A chess chunk is therefore a loose chunk.

In contrast, consider the original example of chunking: recoding binary digits into digits of base 10 (Miller, 1956). This process produces tight chunks. For example, recoding 111_2 as 7_{10} does not retain the three *1* symbols as visible parts of the *7* symbol. Therefore, it is a tight chunk. We do not want to imply that tight chunks are not decomposable. The arabic digit *7* could be further decomposed in a horizontal and a diagonal line. However, those lines would not be meaningful in the context of numbers.

We hypothesize that chunk decomposition begins with loose chunks. Only if the failure persists will the process continue with the decomposition of tight chunks. To predict the relative difficulty of any two problems, one needs to (a) identify which chunks are likely to be active, (b) classify them with respect to tightness, and (c) decide which chunks need to be decomposed to find the relevant solutions. Everything else being equal, a problem solution that requires the decomposition of a tight chunk is more difficult to think of than a problem solution that requires the decomposition of a loose chunk.

Transfer. Once a chunk has been decomposed, it ought to stay decomposed. A single encounter with a particular problem should be sufficient to remove any difficulty associated with the need to decompose chunks that are inappropriate for that problem. Hence, differences in initial difficulty should disappear.

Application

Matchstick Arithmetic

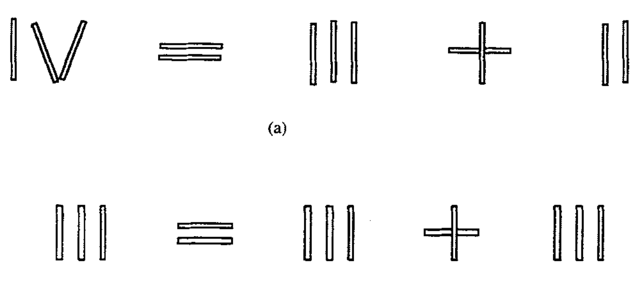
A matchstick arithmetic problem consists of a false arithmetic statement written with Roman numerals (I, II, III, etc.), arithmetic operations (+, −), and equal signs constructed out of matchsticks. Figure 1 shows two example problems in the format in which they were presented to the participants in our experiments. The goal is to move a single stick in such a way that the initial false statement is transformed into a true arithmetic statement. A move consists of grasping a single stick and moving it, rotating it, or sliding it. The rules are that (a) only one stick is to be moved; (b) a stick cannot be discarded, that is, it can be moved only from one position in the equation to another; and (c) the result must be a correct arithmetic statement. An additional rule is that slanted sticks cannot be interpreted as *I* and that the symbols *V* and *X* always consist of two slanted sticks. We refer to the initial and goal statements as "equations" even though they do not contain variables. Each problem has a unique solution.

For example, the false equation in Figure 1A can be transformed into the true equation

$$\mathrm{VI} = \mathrm{III} + \mathrm{III}$$

by removing the left-most stick from *IV* and putting it immediately to the right of the remaining *V*.

All matchstick arithmetic problems considered in this article consist of three roman numerals separated by two arithmetic signs and are solved with a single move. Hence, differences in difficulty are solely a function of how hard it is to think of the right move.



(a)

(b)

Figure 1. Two matchstick arithmetic problems in the format used in the experiments.

Constraints in Matchstick Arithmetic

Problem solvers who are familiar with arithmetic and algebra but who are novices with respect to matchstick arithmetic are likely to assimilate the latter to their knowledge of the former. However, if the constraints of arithmetic and algebra are adhered to, matchstick arithmetic problems are unsolvable (Knoblich & Haider, 1996). The problem solvers must relax one or more of the following three constraints:

1. Value constraint: A numerical value cannot be changed except through operations that produce compensating changes in other values, as when the same quantity is added to or subtracted from both sides of an equation.

2. Operator constraint: An arithmetic function (e.g., addition or subtraction) cannot be arbitrarily deleted, introduced, or altered, except through operations that make corresponding changes elsewhere in the equation. The same is true of the equal sign.

3. Tautology constraint: Arithmetic statements are supposed to have the general form

$$X = f(Y, Z),$$

where f is addition, subtraction, or some other arithmetic function, because their purpose is to specify a calculation to be performed. Tautological statements of the general form $X = X$ are meaningless. (They have their uses in more advanced mathematics, e.g., as starting points for proofs, but not in elementary arithmetic.)

We assume that the visual system parses matchstick arithmetic problems into a representation with three levels: numerals (I, II, III, etc.), functional terms (I + V, III − II, etc.), and entire equations (VI = V + I, III = II − I, etc.). The higher the level at which a change is introduced, the more encompassing is the resulting revision of the representation.

The three constraints map onto the three representational levels as follows: (a) The value constraint applies at the level of numerals; (b) the operator constraint applies at the level of functional terms; and (c) the tautology constraint applies to changes that transform the structure of an entire equation, for example, from a regular to a tautological form, or vice versa.

We hypothesize that constraints are more difficult to relax the higher the level at which they apply. That is, we predict that the tautology constraint has a lower probability of being relaxed than the operator constraint, which in turn has a lower probability of being relaxed than the value constraint.

To solve a matchstick arithmetic problem, one needs to relax one or more constraints, but different problems require relaxation of different constraints. For example, moving a stick from one numeral to another requires that the value constraint be relaxed, whereas changing a plus sign into an equal sign by rotating the vertical stick requires that the operator constraint be relaxed. The mapping between constraints and moves allows one to predict for any two moves which move is easier. By considering which moves are required to solve particular problems, these predictions are transformed into predictions about the relative difficulty of problems (or types of problems).

Chunks in Matchstick Arithmetic

The composite numerals (II, IV, VIII, XI, etc.) are loose chunks in the technical sense introduced in the General Theory section. For example, although the composite numeral *VII* is perceived as a symbol for the number *7,* it is also perceived as consisting of the symbol *V*, followed by two tokens of the symbol *I*. Even though *VII* is a chunk, its parts (*V*, *I*, and *I*) are chunks as well.

In contrast, the three numerals I, V, and X and the minus sign are tight chunks. For example, a *V* sign is perceived as a single unit. There are few contexts (other than matchstick arithmetic) in which it is useful to perceive a *V* or an *X* as consisting of separate lines. Hence, this decomposition has little basis in prior experience. In addition, *V* and *X* do not decompose into meaningful units. Slanted sticks do not have a meaning in matchstick arithmetic.

The plus sign and the equal sign share some features with both loose and tight chunks. On the one hand, they decompose into potentially meaningful components. The plus sign has two perceptually obvious components: a vertical line and a horizontal line. In the context of matchstick arithmetic, the vertical line is equivalent to the symbol *I* for unity, and the horizontal line can be seen as a minus sign. Hence, the plus sign decomposes into two meaningful chunks. The equal sign decomposes into two perceptually obvious components, one of which can be interpreted as a minus sign. (It is unclear how the second component might be interpreted in this case.) On the other hand, neither plus signs nor equal signs are commonly analyzed in this way, so these decompositions have little support from prior experience. Hence, we tentatively assigned the plus and equal signs to a category of "intermediate chunks."

Different matchstick arithmetic moves require decomposition of different chunks. For example, moving a stick from a symbol like *VII* requires that the chunk for *VII* is decomposed into its components, *V*, *I*, and *I*. Similarly, changing a plus sign into an equal sign by rotating the vertical stick requires that the source chunk "+" is decomposed into its two matchstick components, "−" and "|," and that the target chunk "=" is decomposed into its components. We hypothesize that tight chunks have a lower probability of being decomposed than intermediate chunks, which in turn have a lower probability of being decomposed than loose chunks.

By mapping moves onto chunk types, one can predict, for any two moves, which move is easier. For example, moving the numeral *I* from one composite numeral to another respects the integrity of all tight chunks. Moving a matchstick from a tight chunk to a composite numeral or vice versa should be more difficult. A move that removes a matchstick from a tight chunk, changes its orientation or location, and adds it again to the same chunk, thus changing the symbol into another, is a special case. Such a move requires the decomposition of a single numeral, but it requires the decomposition of two chunk types: the source chunk and the target chunk.

By considering which moves are required to solve particular problems, one can predict the relative difficulty of individual matchstick arithmetic problems. Everything else being equal, the tighter the chunks that need to be decomposed, the lower the probability of thinking of the corresponding move. By considering which move is required to solve a particular problem, one can convert such predictions into predictions about the relative difficulty of particular problems (or problem types).

Problem Types

The purpose of this section is to define four types of matchstick arithmetic problems and to rank order them with respect to difficulty. We describe each type in terms of the constraints that need to be relaxed and the tightness of the chunks that need to be decomposed.

1. Type A (value constraint; loose chunks). The problem

$$\mathrm{VI} = \mathrm{VII} + \mathrm{I}$$

is solved by moving one matchstick from the numeral *VII* to the numeral *VI* to produce the equation

$$\mathrm{VII} = \mathrm{VI} + \mathrm{I}.$$

This move requires relaxation of the value constraint (but no other constraint). In addition, it requires decomposition of the source chunk *VI* and the target chunk *VII*, both of which are loose chunks.

2. Type B (value and operator constraints; loose and intermediate chunks). The solution to the problem

$$\mathrm{I} = \mathrm{II} + \mathrm{II}$$

is to remove one stick from the plus sign, thereby changing addition into subtraction, and add that stick to the first term to the right of the equal sign. The result is

$$\mathrm{I} = \mathrm{III} - \mathrm{II}.$$

This move violates the value and operator constraints. In addition, it requires decomposition of the plus sign (the source chunk) and the numeral *III* (the target chunk). Hence, problems of Type B should be more difficult than problems of Type A.

3. Type C (operator and tautology constraints; intermediate chunks). The solution to

$$\mathrm{III} = \mathrm{III} + \mathrm{III}$$

is to remove the arithmetic operation by rotating the vertical stick of the plus sign, producing the following solution:

$$\mathrm{III} = \mathrm{III} = \mathrm{III}.$$

This solution violates both the operator and tautology constraints. In addition, it requires decomposition of both the plus sign and the equal sign. Hence, problems of Type C should be more difficult than problems of Type B. Problem Types A through C form a ladder where the constraints become wider or the chunks tighter for each rung.

4. Type D (value constraint; tight chunks). The solution to

$$\mathrm{XI} = \mathrm{III} + \mathrm{III}$$

is to slide one of the matchsticks that make up the symbol *X* so as to change the latter into the symbol *V*, giving the result

$$\mathrm{VI} = \mathrm{III} + \mathrm{III}.$$

This move violates the value constraint but no other constraint. However, it requires that the two tight chunks *X* and *V* are decomposed into their components. Hence, problems of Type D should be more difficult than problems of Type A. Without an interval scale for measuring difficulty, we could not predict the relative difficulty of Type D versus Type B or C. The purpose of including Type D problems in the study was to ascertain the effect of chunk decomposition independently of constraint relaxation by comparing performance on Type D problems with performance on Type A problems.

Transfer

A single encounter with a matchstick problem of a given type should be sufficient to remove the sources of difficulty associated with that problem type. Hence, there should be no significant differences in difficulty among the four problem types on a second encounter. Specifically, removing sources of difficulty should reduce the difficulty of Problem Types B, C, and D to the difficulty level of Problem Type A, the simplest problem type. This argument predicts an interaction between initial difficulty and transfer, with the simplest problem type (Type A) showing the least amount of transfer and the most difficult (Types C and D) the largest.

Experiment 1A

The purpose of Experiment 1A was to verify that constraint relaxation and chunk decomposition influence performance on matchstick arithmetic problems. The need for both processes was systematically varied across problems in a within-participant design. The participants solved two blocks of problems. The problems within each block exemplified the four different Problem Types A through D. The problems in the second block were matched to the problems in the first block with respect to type. The following hypotheses were tested:

1. Constraint relaxation hypothesis: The probability of relaxing a constraint is lower the wider the scope of the constraint. Given a limit on the amount of time available to find a solution, the frequency of solution is highest for Type A, lower for Type B, and lowest for Type C on the first block of problems. Solution times in Block 1 are highest for Type C, lower for Type B, and still lower for Type A.

Table 1
Matchstick Arithmetic Problems Used in Experiment 1A by Type and Block

	Block 1		Block 2	
Type	Problem	Solution	Problem	Solution
A	VI = VII + I	VII = VI + I	II = III + I	III = II + I
A	IV = III + III	VI = III + III	IX = VIII + III	XI = VIII + III
B	I = II + II	I = III − II	III = V + III	III = VI − III
B	IV = III − I	IV − III = I	V = III − II	V − III = II
C	III = III + III	III = III = III	IV = IV + IV	IV = IV = IV
D	XI = III + III	VI = III + III	VI = VIII + III	XI = VIII + III

2. Chunk decomposition hypothesis: The probability of decomposing a chunk is lower if tight chunks have to be decomposed than if only loose chunks have to be decomposed. Given a time limit, the frequency of solution should be higher for Type A than for Type D on the first block of problems. Solution times in the first block should be higher for Type D problems.

3. Transfer hypothesis: Once relaxed, constraints stay relaxed; once decomposed, tight chunks stay decomposed. That is, there are no significant differences with respect to frequency of solution and solution time among the four Problem Types A through D on the second block of problems. Furthermore, there is an interaction between problem type and encounter, such that performance on more difficult problem types changes more from the first to the second block of problems.

Method

Participants. Twenty undergraduates (13 women and 7 men) from the University of Hamburg participated in the study for course credit. They ranged in age from 19 to 45 years. We selected only participants who were familiar with Roman numerals. Because there were only within-participant treatments in this experiment, all participants were assigned to the same condition. The participants were seen individually.

Procedure, materials, and design. The participants first went through a training phase aimed at reducing individual differences in how fast they could recognize Roman numerals. Each participant looked at numerals presented on a computer screen, pushed a button as soon as he or she had identified the value of the presented numeral, and then said the value out loud. No feedback was given. There were five training blocks, with 75 trials per block. Within each block, all Roman numerals *I* and *XV* were presented five times each in random order.

In the experimental phase, there were two blocks of six matchstick arithmetic problems. Each block contained two instances of Type A, two instances of Type B, one instance of Type C, and one instance of Type D (see Table 1). The six problems were presented in random order within blocks. A new random order was generated by the computer for each participant.

The problems were presented on a computer screen. Participants pushed a button as soon as they knew the correct solution and said it out aloud afterward. If the solution was correct, they were presented with the next task. If the solution was not correct, they continued to work on the same problem until they found another solution or until they reached the time limit of 5 min. If participants did not solve a problem within 5 min, the trial was interrupted, and they were told the solution.

Results

The dependent variables of this experiment were the frequency of problems solved within the 5-min limit and solution times. The results are reported separately for each variable.

Solution frequencies. The results showed the predicted order of task difficulty for the first block of problems. Problems that required the relaxation of constraints with wider scope were solved less frequently (Type C < Type B < Type A). Figure 2 shows the cumulative frequencies of solutions for Types A through C. The predicted ordering of the problem types held at each interval. Problems that required decomposition of tight chunks were solved less frequently than problems that required only decomposition of loose chunks (Type D < Type A). Figure 3 shows the cumulative frequencies of solutions for Types A and D. The predicted rank ordering held at each 1-min interval.

There were no significant differences in difficulty among

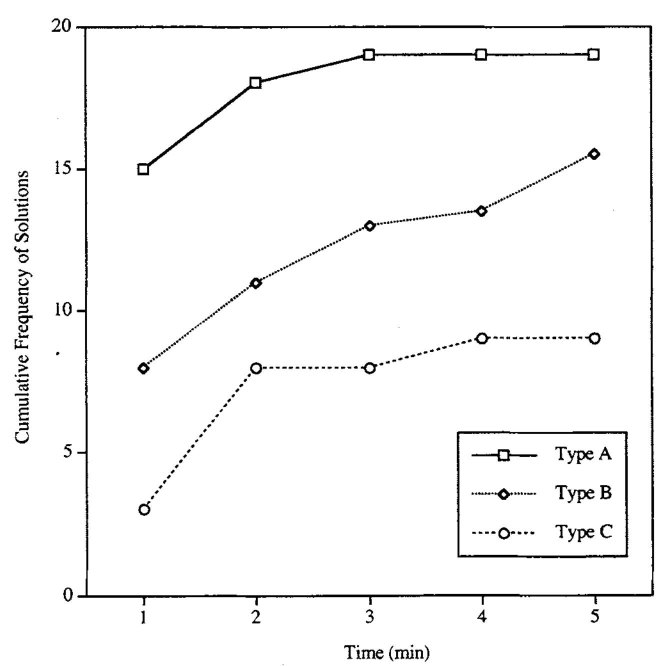


Figure 2. Cumulative solution rates for Problem Types A–C in Block 1 of Experiment 1A.

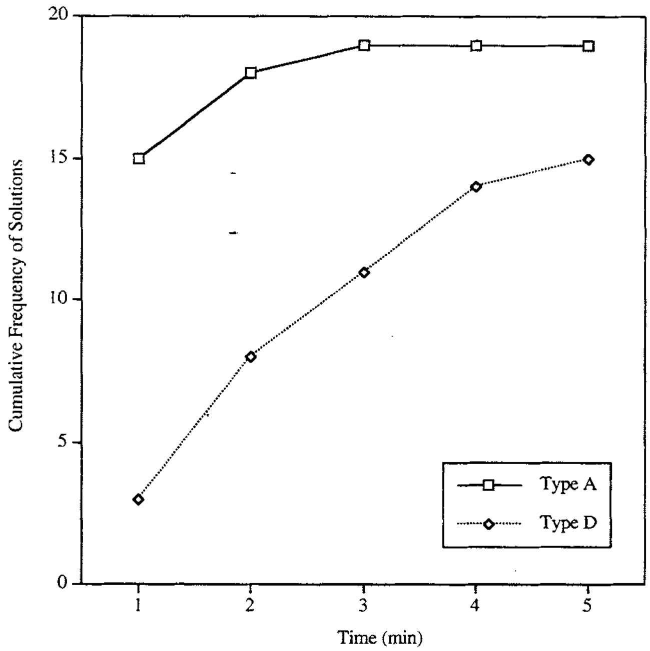


Figure 3. Cumulative solution rates for Problem Types A and D in Block 1 of Experiment 1A.

the four problem types in the second block of problems. Figure 4 shows the cumulative frequencies of solutions on the second block across 1-min intervals for each problem type. The four curves are almost superimposed. In addition,

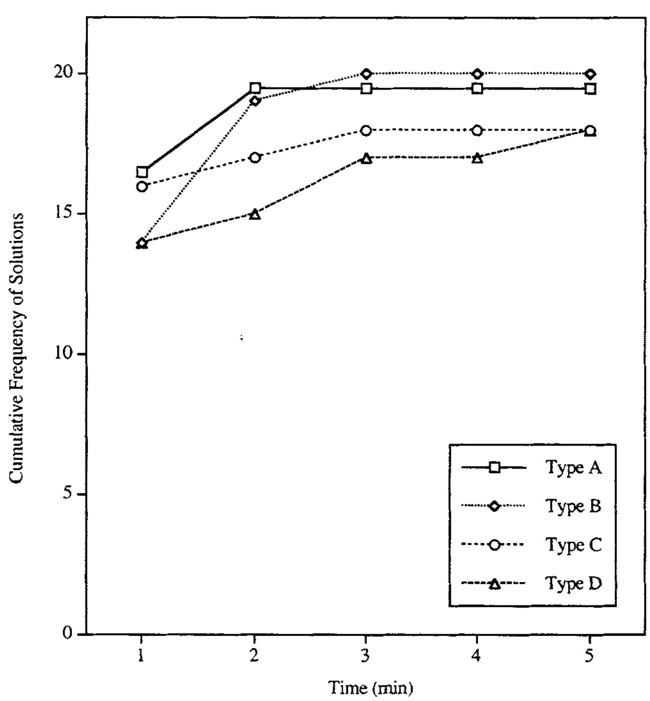


Figure 4. Cumulative solution rates for Problem Types A–D in Block 2 of Experiment 1A.

the contrast between Figures 2 and 3, on the one hand, and Figure 4, on the other hand, shows the predicted interaction between initial difficulty and transfer: The upward displacement of the curves in Figure 4 as compared with Figures 2 and 3 is larger the more difficult the problem type.

An alpha level of .05 was used for all statistical tests. A 4 × 2 analysis of variance (ANOVA) with the variables problem type (A, B, C, or D; within participants) and block (1 vs. 2; within participants) was computed on solution frequency. The main effect of problem type was significant, $F(3, 57) = 7.11$, $MSE = 0.0838$, $p < .001$. Post hoc comparisons revealed that significantly fewer problems of Type B were solved than of Type A ($p < .05$) and significantly fewer problems of Type C than of Type B ($p < .01$). A further comparison revealed that problems of Type D were solved less frequently than problems of Type A ($p < .05$). In summary, the differences were C < B < A and A < D.

Furthermore, there was a significant main effect of block, $F(1, 19) = 14.41$, $MSE = 0.1253$, $p < .01$. More problems were solved during Block 2. There were no significant differences among problem types during the second block (all ps > .10).

Consequently, there was a significant two-way interaction between problem type and block, $F(3, 57) = 3.24$, $MSE = 0.0984$, $p < .05$. The amount of transfer between blocks differed for different problem types. Post hoc comparisons revealed a significant transfer effect for those problems that were solved less frequently during the first block, namely, Type B ($p < .05$) and Type C ($p < .001$). No significant transfer was observed for problems of Type A ($p = .80$). The transfer effect for problems of Type D (75% solved on Block 1 vs. 90% solved on Block 2) did not reach statistical significance ($p = .13$).

Solution times. In a next step, solution times for problems of Types A through D were analyzed. For problems that were not solved, the solution time was replaced with the upper time limit of 5 min. The resulting measure was a conservative estimate of the actual solution time because the solution time was underestimated for problems that were not solved.

Because the distribution of solution times was highly skewed, we report medians and quartiles and used ordinal tests to assess statistical significance. Table 2 shows medians and lower and upper quartiles for problems solved in Block 1. The medians were consistent with the rank ordering predicted on the basis of the scope of the relevant constraints (Type A < Type B < Type C) as well as the rank ordering

Table 2
Solution Times (in Seconds) for Block 1 of Experiment 1A (N = 20)

	Problem type			
Statistic	A	B	C	D
Mdn	40	155	300	140
Lower *Q*	14	67	91	98
Upper *Q*	61	190	300	290

predicted on the basis of tightness of the chunks (Type A < Type D).

We computed single Wilcoxon tests to assess the statistical significance of the differences observed in Block 1. Problems of Type A were solved significantly faster than problems of Type B, $N = 20$, $T = 12$, $p < .001$, and problems of Type B were solved significantly faster than problems of Type C, $N = 20$, $T = 41$, $p < .05$. Moreover, problems of Type A were solved significantly faster than problems of Type D, $N = 20$, $T = 10$, $p < .001$. All differences predicted by the constraint relaxation and chunk decomposition hypotheses were statistically significant.

The solution times for Block 2 did not exhibit the rank ordering observed in Block 1 (see Table 3). Single Wilcoxon tests revealed no significant differences between problems of Types A, B, and C and problems of Types A and D (all $ps > .10$).

To test our transfer predictions, we compared solution times in the first and second blocks for each problem type. Single Wilcoxon tests showed significant transfer for problems of Type B, $N = 20$, $T = 14$, $p < .001$; Type C, $N = 20$, $T = 23$, $p < .01$; and Type D, $N = 20$, $T = 32$, $p < .05$, but not for problems of Type A, $N = 20$, $T = 102$, $p = .91$.

In short, problems that required the relaxation of wider constraints took longer to solve, as did problems that required the decomposition of tight chunks. After the first encounter with the relevant problem types, these differences disappeared. Amount of transfer from a first to a second encounter was a function of initial difficulty, with more difficult problem types exhibiting larger transfer effects.

Discussion

The results of Experiment 1A are consistent with the hypotheses that constraint relaxation and chunk decomposition affect performance on matchstick arithmetic problems. The frequency and speed of solution depend on which, and how many, constraints have to be relaxed to find the solution and whether the perceptual chunks that have to be decomposed are loose, intermediate, or tight, in the technical sense defined in the General Theory section.

The results verify that it is easier to realize that one can arbitrarily change values in matchstick arithmetic than to realize that one also can change the operators, which in turn is easier than to realize that one can transform an equation into a tautology. It is easier to decompose multidigit roman numerals into their component numerals than to decompose plus signs, minus signs, equal signs, and single-digit roman numerals. The fact that tight chunks add an increment of difficulty independently of constraint relaxation (Type D is more difficult than Type A) supports the hypothesis that we have identified two separate sources of difficulty in matchstick arithmetic problems.

Without an interval scale for measuring amount of difficulty, we could not predict the difficulty of Type D problems in relation to Types B and C because we had no way of calculating the trade-off between the need for constraint relaxation and the need for chunk decomposition. The data show that problems of Type D were approximately as difficult as problems of Type B.

The strength of the transfer effect observed in Experiment 1A is noteworthy. A single exposure to a problem of Type B, C, or D was sufficient to make those problems as simple as problems of Type A. The only deviation from our predictions was that transfer was not significant for Type D problems in the analysis of solution frequency. However, the analysis of solution times showed that Type D problems became significantly easier when encountered for the second time.

Table 3
Solution Times (in Seconds) for Block 2 of Experiment 1A ($N = 20$)

Statistic	Problem type A	B	C	D
Mdn	33	38	10	42
Lower *Q*	18	27	7	20
Upper *Q*	59	73	47	124

Experiment 1B

In Experiment 1A, participants solved only one instance of each problem type. Hence, the observed differences in task difficulty may not have been due to the scope of constraints and the tightness of chunks as proposed in the theoretical analysis but rather to material effects related to the specific tasks used in Experiment 1A. We conducted another experiment to support the generality of the problem type effect by assessing the relative difficulty of two additional problems of each type.

Method

Participants. The 38 paid participants, 17 male and 21 female, were recruited by advertising at the University of Munich campus and in the local newspapers. They ranged in age from 20 to 36 years. They were randomly assigned to two experimental conditions. Participants in the first condition solved Version 1 of each problem type; participants in the second condition solved Version 2 of each problem type.

Procedure, materials, and design. The procedure was the same as that used in Experiment 1A with the exception that participants solved only one block of matchstick arithmetic problems. Table 4 shows the two versions of each problem type that were used in Experiment 1B. As in Experiment 1A, the dependent variables were the frequency of problems solved within the 5-min limit and the solution times.

Results

Solution frequencies. The results showed the same pattern as that in Experiment 1A. Problems that required the relaxation of constraints with wider scope were solved less frequently (Type C < Type B < Type A). Figure 5 shows the cumulative frequencies of solutions across 1-min intervals for problems of Types A through C collapsed over both experimental conditions. The predicted ordering of the problem types held at each interval. Furthermore, problems

Table 4
Problems Used in Experiment 1B

	Version 1		Version 2	
Type	Problem	Solution	Problem	Solution
A	IV = II + IV	VI = II + IV	III = IV − III	III = VI − III
A	VII = II + III	VI = III + III	XI = I + XII	XII = I + XI
B	IV = VI + I	IV = VI − II	II = V + IV	II = VI − IV
B	VIII = VI − II	VIII − VI = II	VI = IV − II	VI − IV = II
C	IV = IV + IV	IV = IV = IV	VI = VI + VI	VI = VI = VI
D	IV = III + VI	IX = III + VI	IV = V + IV	IX = V + IV

that required the decomposition of tight chunks were solved less often than problems that required the decomposition of loose chunks (Type D < Type A). Figure 6 shows the cumulative frequencies of solutions across 1-min intervals for problems of Types A and D collapsed over both experimental conditions. The predicted ordering of the problem types held at each interval.

A 2 × 4 ANOVA with the variables experimental condition (Version 1 vs. Version 2; between participants) and problem type (A, B, C, or D; within participants) was computed on solution frequency. There was no main effect for experimental condition, $F(1, 36) = 0.39$, $MSE = 0.1503$, $p = .52$; that is, there were no overall differences in difficulty between the two problem versions. Post hoc comparisons revealed that there was no significant difference between the two versions in any of the problem types (all $ps > .10$). However, the main effect of problem type was significant, $F(4, 144) = 10.03$, $MSE = 0.1533$, $p < .001$. Post hoc comparisons showed that significantly fewer problems of Type C were solved than problems of Type B ($p < .001$). Although fewer problems of Type B were solved than problems of Type A, the predicted difference failed to reach statistical significance ($p = .12$). A further comparison revealed that problems of Type D were solved less frequently than problems of Type A ($p < .05$). The pattern of results is almost the same as that in Experiment 1A, with the differences being C < B ≤ A and A < D.

Solution times. In a second step, solution times for problems of Types A through D were analyzed. For problems that were not solved, the solution time was replaced with the upper time limit of 5 min. Again, the distribution of solution times was highly skewed. Therefore, we report medians and quartiles and used ordinal tests to assess statistical significance.

Table 5 shows that the medians and upper quartiles of each problem type followed the rank ordering predicted by the scope of constraints (Type A < Type B < Type C). The medians also followed the rank ordering predicted by the chunk decomposition hypothesis (Type A < Type D).

We conducted single U tests to detect differences in solution times between the two versions of each problem

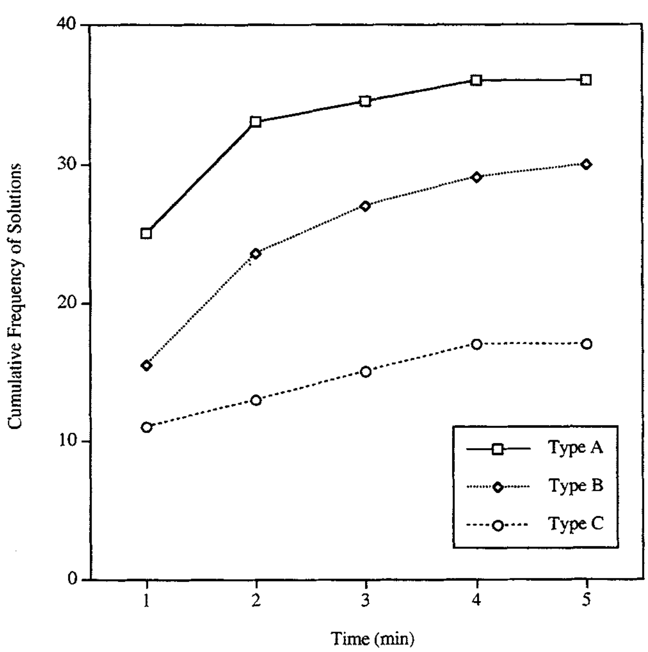


Figure 5. Cumulative solution rates for Problem Types A–C in Experiment 1B.

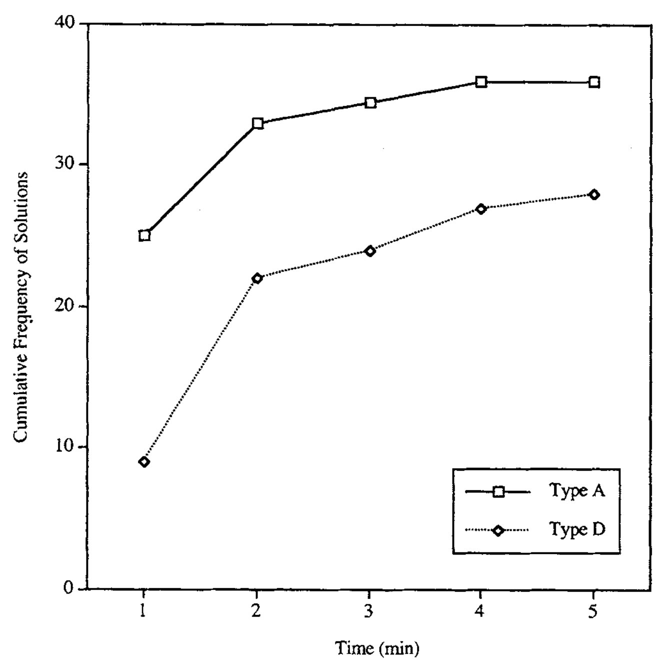


Figure 6. Cumulative solution rates for Problem Types A and D in Experiment 1B.

Table 5
Solution Times (in Seconds) in Experiment 1B

Statistic	Problem type A	B	C	D
Mdn	43	126	300	104
Lower *Q*	23	57	30	63
Upper *Q*	76	165	300	300

type. None of them showed a significant difference (all $ps > .10$). We computed single Wilcoxon tests to assess if the differences between problem types were statistically significant. Problems of Type A were solved significantly faster than problems of Type B, $N = 38$, $T = 107$, $p < .001$, and problems of Type B were solved significantly faster than problems of Type C, $N = 38$, $T = 160$, $p < .01$. Moreover, Type A problems were solved significantly faster than Type D problems, $N = 38$, $T = 96$, $p < .001$. As in Experiment 1A, problems that required relaxation of wider constraints and problems that required decomposition of tight chunks took longer to solve.

Discussion

The results of Experiment 1B show that the main problem type effect reported in Experiment 1A does not depend on specific material effects. The difficulty of two more versions of each problem type followed the predicted pattern. The frequency of solution depends on the scope of constraints that must be relaxed to find a solution and the tightness of perceptual chunks to be decomposed. No differences were observed between experimental groups that solved different versions of each problem type. The results of Experiment 1B support the categorization of problem types according to the constraint relaxation and chunk decomposition hypotheses.

The only deviation from our predictions was the lack of a significant difference in solution frequency between problems of Type B and Type A. This nonsignificant difference is probably due to a ceiling effect. The higher the likelihood that a certain problem type is solved within the given time interval, the higher the likelihood of obtaining nonsignificant effects in solution frequency. The huge and highly reliable difference in solution times between problems of Types A and B shows clearly that the predicted difference in task difficulty was present in Experiment 1B. Having shown that the main problem type effect is not due to the specific tasks used in Experiment 1A, we examined the large transfer effect obtained in that experiment.

Experiment 2

Historically, the term *transfer* has been used to refer to improvement in a person's performance on Task B by prior exposure to Task A. The key issue always has been what properties of Tasks A and B mediate transfer effects. Thorndike (1903, 1913) proposed the first systematic theory of transfer. His identical elements hypothesis claimed that the amount of facilitation is a function of the similarity between the two tasks. Singley and Anderson (1989) reformulated the identical elements hypothesis in terms of production rules.

The hypothesis that a problem representation can be changed by constraint relaxation and chunk decomposition implies that those changes should affect performance on other problems in which the same constraints and chunks are relevant. Therefore, the concept of representational change adds an additional source of transfer, over and above problem similarity. Once a problem representation has changed in a particular way, the change persists and so transfers to subsequent problems for which that representation is relevant. In particular, relaxing a constraint affects performance on any task for which that constraint would have blocked the correct solution, had it remained active. For example, relaxing the constraint against arbitrarily changing an operator in an equation has the potential to affect performance on any task for which some operator plays an important role. Similarly, decomposing a tight chunk has the potential to affect performance on any task for which that chunk is relevant. For example, decomposing a number symbol like *X* into its component symbols \ and / has the potential to affect performance on any problem that involves that symbol.

The main purpose of Experiment 2 was to investigate the relative contributions of similarity and representational change to the transfer effects observed in Experiment 1A. To separate these effects, we varied surface similarity orthogonally to the need to relax constraints and decompose chunks.

We varied similarity along two dimensions: the order of the initial equation (i.e., the sequence of the symbols, the location of the equal sign and the function term) and the specific numerals that appeared in that equation. The matchstick arithmetic domain allows us to vary both order and numerals orthogonally to each other. For example, consider the Type A problem

$$\text{IV} = \text{V} - \text{III}, \qquad \text{(Problem 1)}$$

which has the solution IV = VI − II. It can be solved by moving one matchstick from the symbol *III* to the symbol *V*. This move violates the value constraint (but no other constraint) and requires the decomposition of loose chunks (but no tight chunks). A second encounter with Problem 1 obviously maximizes similarity of both numerals and order. We refer to a second presentation of a previously encountered problem as a "same–same problem."

The Type A problem

$$\text{V} - \text{III} = \text{IV} \qquad \text{(Problem 2)}$$

consists of the same numerals as Problem 1, but they are presented in a different order. We refer to Problem 2 as a "same-numerals problem" (with respect to Problem 1).

The Type A problem

$$\text{IX} = \text{X} - \text{III} \qquad \text{(Problem 3)}$$

contains different numerals than Problem 1, but the order is similar. We refer to Problem 3 as a "same-order problem" (with respect to Problem 1).

Finally, consider the Type A problem

$$X - III = IX, \qquad \text{(Problem 4)}$$

which contains different numerals in a different order than those in Problem 1. We refer to Problem 4 as a "different–different problem" (with respect to Problem 1).

The matchstick arithmetic domain allows us to vary both dimensions of similarity within each problem type. For each of the four Problem Types A through D (see the Application section), we can construct problem sets that vary both order and numerals in the way illustrated by Problems 1 through 4 (see the *Method* section of this experiment).

If similarity is the main determinant of the transfer effects observed in Experiments 1A and 1B, then the transfer effects should be smallest for same–same problems, intermediate in size for same-numerals and same-order problems, and largest for different–different problems, regardless of problem type. In contrast, if constraint relaxation and chunk decomposition are the main determinants of the observed transfer effects, then those effects should vary with problem type but be equal across levels of similarity.

In Experiment 2, we attempted a modest generalization to an American sample. Because Roman numerals are common in Germany but rare in the United States, we expected different levels of familiarity with this type of numeral. The pattern of effects observed in Experiments 1A and 1B should be robust with respect to this variable.

Method

Participants. Seventy-eight undergraduates (45 women and 33 men) from the University of Illinois at Chicago participated in the study for course credit. They ranged in age from 18 to 38 years. They were randomly assigned to the four experimental conditions.

Procedure. The training phase of Experiment 2 was identical in purpose and procedure to the training phase of Experiment 1A. As expected, the participants in Experiment 2 (American students) had less prior familiarity with Roman numerals than their counterparts in Experiment 1A (German students). The average time to complete a training trial was 561 ms in Experiment 2 as compared with 457 ms in Experiment 1A.

In the experimental phase, participants encountered two blocks of matchstick arithmetic problems. Each block contained two instances of Type A, two instances of Type B, one instance of Type C, and one instance of Type D. The six problems in Block 1 were identical in the four experimental conditions. These six problems appear in the second column in Table 6.

The four experimental conditions differed in how the six problems in Block 2 related to the problems in Block 1. In the same–same condition, the participants solved the same six problems as those in Block 1. In the same-numerals condition, the participants solved six problems that contained the same numerals as the problems in Block 1 but presented in a different order. These problems are shown in the third column in Table 6. In the same-order condition, the participants solved six problems that presented different numerals in the same order as the problems in Block 1. These problems are shown in the fourth column in Table 6. Finally, in the different–different condition, the participants solved problems that presented different numerals in different orders. These problems are displayed in the fifth column in Table 6.

Problems were presented in random order within blocks. A new random order was generated by the computer for each participant. The problems were presented on a computer screen. Participants pushed a button as soon as they knew the correct solution and then said it out loud. If the solution was correct, they worked on the next problem. If the solution was not correct, they continued to work on the same problem until they announced another solution or encountered the 5-min time limit. If participants did not solve a problem within 5 min, they were told the solution.

Results

The frequency of solution and the time to solution were the dependent variables. The results from each variable are reported separately.

Table 6
Problems Used in Experiment 2 by Type and Condition

Problem type	Experimental condition: Same–same	Same-numerals	Same-order	Different–different
A				
Problem	IV = V − III	V − III = IV	IX = X − III	X − III = IX
Solution	IV = VI − II	VI − II = IV	IX = XI − II	XI − II = IX
A				
Problem	IV = III + III	III + III = IV	IV = II + IV	II + IV = IV
Solution	VI = III + III	III + III = VI	VI = II + IV	II + IV = VI
B				
Problem	I = II + II	II + II = I	III = V + III	V + III = III
Solution	I = III − II	III − II = I	III = VI − III	VI − III = III
B				
Problem	IV = III − I	I − IV = III	VIII = V − III	III − VIII = V
Solution	IV − III = I	I = IV − III	VIII − V = III	III = VIII − V
C				
Problem	III = III + III	III + III = III	IV = IV + IV	IV + IV = IV
Solution	III = III = III	III = III = III	IV = IV = IV	IV = IV = IV
D				
Problem	XI = III + III	III + III = XI	IX = I + III	I + III = IX
Solution	VI = III + III	III + III = VI	IV = I + III	I + III = IV

Table 7
Percentage Solutions in Experiment 2 by Condition and Block

Condition	Block 1	2	M
Same–same	67	96	82
Same-numerals	64	99	82
Same-order	64	95	80
Different–different	71	95	83
M	66	96	82

Solution frequencies. Table 7 shows the frequency of solution for the four similarity conditions. There were minimal differences across conditions. Hence, similarity appears not to have been a strong determinant of transfer effects in this task domain.

However, the results for Block 1 showed the order of task difficulty predicted by the constraint relaxation and chunk decomposition hypotheses. Fewer problems were solved within the time limit if the constraints to be relaxed had a wider scope (Type C < Type B < Type A). Figure 7 shows the cumulative frequencies of solutions for 1-min intervals for Types A through C. The predicted rank ordering of the problem types held at each interval.

Problems that required decomposition of tight chunks were solved less often than problems that required decomposition of loose chunks only (Type D < Type A). Figure 8 compares the cumulative solution frequencies for 1-min intervals for Types D and A. The differences were in the predicted direction at each interval.

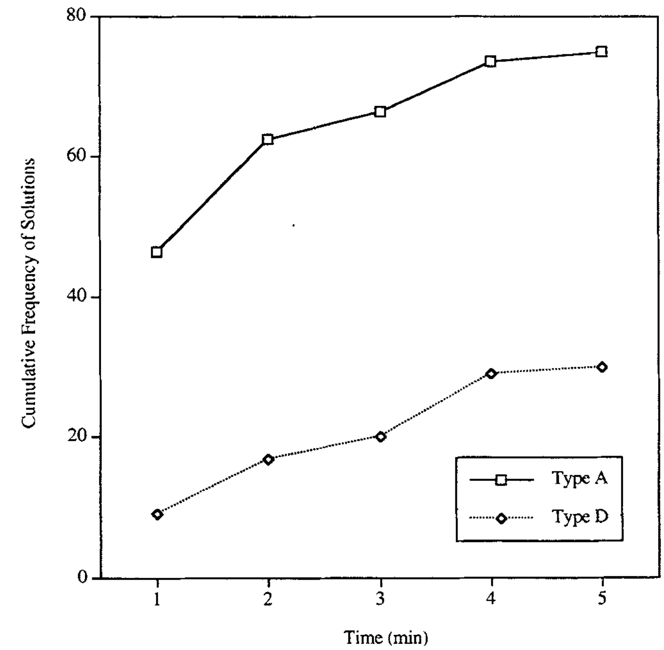


Figure 8. Cumulative solution rates for Problem Types A and D in Block 1 of Experiment 2.

There was a transfer effect from Block 1 to Block 2. The size of the effect did not differ among the four experimental conditions. However, the transfer effect was larger for more difficult problem types. Figure 9 shows the cumulative frequency of solution for all four problem types on the second block, aggregated across the four experimental

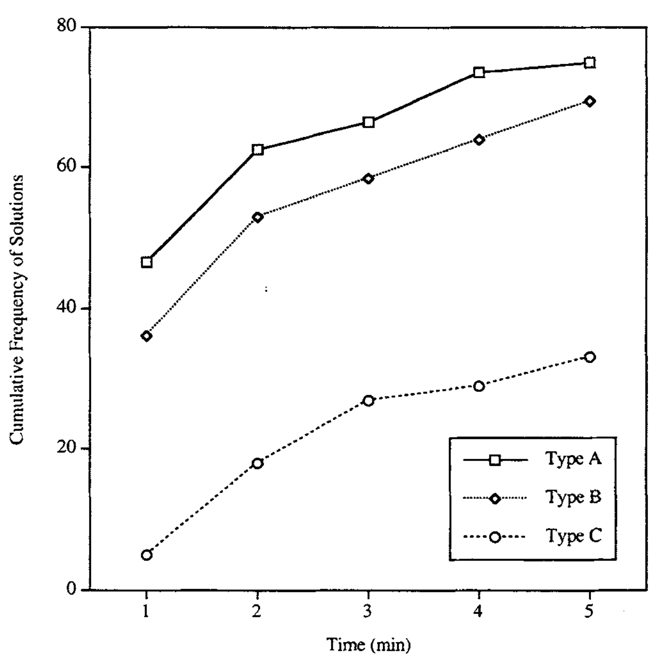


Figure 7. Cumulative solution rates for Problem Types A–C in Block 1 of Experiment 2.

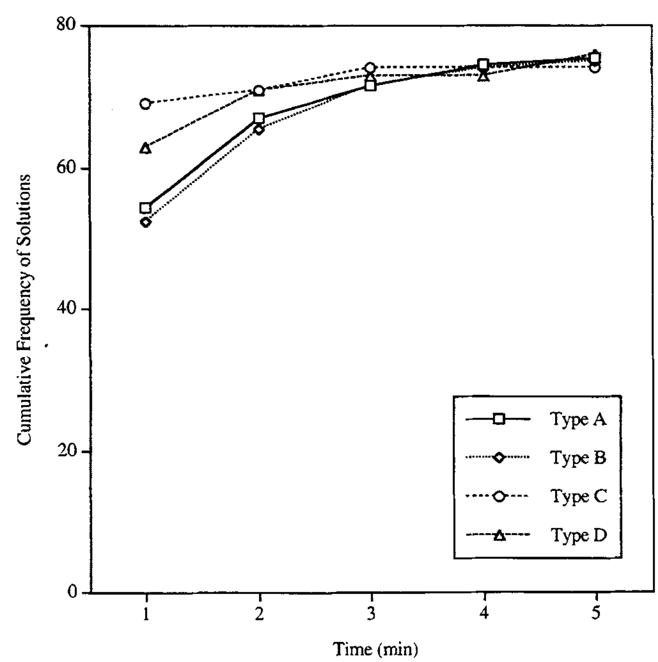


Figure 9. Cumulative solution rates for Problem Types A–D in Block 2 of Experiment 2.

conditions. The relative lack of differences between problem types in Block 2 is reflected in the overlap between the curves. The larger transfer effects for more difficult problem types are reflected in the upward shifts of the curves for Types B, C, and D (compared with the corresponding curves in Figures 7 and 8) and the relative lack of such a shift for Type A.

An alpha level of .05 was used for all statistical tests. A 4 × 4 × 2 ANOVA with the variables similarity (same–same, same-numerals, same-order, or different–different; between participants), problem type (A, B, C, or D; within participants), and block (1 vs. 2; within participants) was computed on solution frequency. There was no significant main effect for similarity, $F(3, 74) = 0.34$, $MSE = 0.0791$, $p = .79$.

The main effect of problem type was significant, $F(4, 222) = 46.45$, $MSE = 0.0782$, $p < .001$; that is, some problem types were more difficult than others. Post hoc comparisons between problem types in Block 1 revealed that the predicted difference between problems of Type A and Type B failed to reach statistical significance ($p = .14$). However, problems of Type B were solved significantly less often than problems of Type C ($p < .001$). A further post hoc comparison revealed that problems of Type A were solved significantly less often than problems of Type D ($p < .001$).

There was a significant main effect of the block variable, $F(1, 74) = 177.97$, $MSE = 0.0778$, $p < .001$. More problems were solved during the second block. Post hoc comparisons revealed significant transfer for problems of Types C ($p < .001$) and D ($p < .001$), that is, problems that required relaxation of the most fundamental constraint and problems that required decomposition of tight chunks. No significant transfer was observed in problems of Type A ($p = .89$). Contrary to our expectations, no significant transfer was observed for problems of Type B ($p = .14$).

As these results imply, there was a significant two-way interaction between problem type and block, $F(4, 222) = 39.99$, $MSE = 0.0888$, $p < .001$. Problem types that were more difficult in the first block were solved more often in the second block than in the first block. Consequently, there were no significant differences in solution frequency between problem types in the second block (all $ps > .10$); all problems were reduced to the difficulty level of the simplest problems. The two-way interactions between similarity and problem type and between similarity and block and the three-way interaction between similarity, problem type, and block were not significant.

In a further post hoc comparison, we compared the difficulty of problem types in the second block in the same–same and different–different conditions. We found no significant differences between these two conditions for Problem Type A ($p = .98$), B ($p = .60$), C ($p = .30$), or D ($p = .27$).

Solution times. The distribution of solution times was highly skewed. Therefore, we report medians and quartiles and used ordinal tests to assess statistical significance. First, we analyzed the main effect of task difficulty in the first block of problems, and then we analyzed the transfer effects.

Table 8
Solution Times (in Seconds) for Block 1 of Experiment 2

	Problem type			
Statistic	A	B	C	D
Mdn	61	95	300	300
Lower *Q*	28	46	126	157
Upper *Q*	116	166	300	300

Table 8 shows medians and lower and upper quartiles of the solution times for problems solved in Block 1. The medians exhibited the predicted rank ordering (Type A < Type B < Type C). The solution times were longer for problems of Type D than for problems of Type A.

We computed single Wilcoxon tests to assess if the observed differences were statistically reliable. Problems of Type A were solved significantly slower than problems of Type B, $N = 78$, $T = 913$, $p < .01$; problems of Type B significantly slower than problems of Type C, $N = 78$, $T = 264$, $p < .001$; and problems of Type D significantly slower than problems of Type A, $N = 78$, $T = 151$, $p < .001$. These results replicate those of Experiments 1A and 1B. Solution times were higher for problems that required relaxation of wider constraints and problems that required decomposition of tight chunks.

The rank ordering of solution times for problems of Types A through D observed in Block 1 did not persist in Block 2 (see last three rows of Table 9). Unexpectedly, there were significant differences between problem types in Block 2. Wilcoxon tests showed that Type C problems were solved significantly faster than problems of Type A, $N = 78$, $T = 595$, $p < .001$, and Type B, $N = 78$, $T = 581$, $p < .001$. Furthermore, problems of Type D were solved faster than problems of Type A, $N = 78$, $T = 1,008$, $p < .01$.

Table 9
Solution Times (in Seconds) for Block 2 of Experiment 2

	Problem type			
Condition and statistic	A	B	C	D
Same–same ($n = 19$)				
Mdn	47	39	6	16
Lower *Q*	12	30	5	7
Upper *Q*	84	60	8	62
Same-numerals ($n = 20$)				
Mdn	40	43	8	10
Lower *Q*	31	23	5	8
Upper *Q*	78	71	16	27
Same-order ($n = 19$)				
Mdn	44	50	8	26
Lower *Q*	27	33	6	13
Upper *Q*	57	139	14	73
Different–different ($n = 20$)				
Mdn	54	57	19	17
Lower *Q*	26	39	8	8
Upper *Q*	118	98	40	27
All ($N = 78$)				
Mdn	45	48	8	18
Lower *Q*	26	32	5	8
Upper *Q*	84	94	18	37

These differences were not due to the persistence of the differences observed in Block 1. The statistically significant differences in Block 2 reversed the direction of the differences in Block 1. Therefore, the differences were unexpected but not inconsistent with our theory.

The lack of differences between the experimental groups in the analysis of solution frequencies may be partly due to ceiling effects. To address this problem, we conducted a further analysis of solution times in the transfer block, which provide a more sensitive measure to detect effects of similarity on transfer. Table 9 shows the solution times for each problem type in each group.

We conducted Kruskal–Wallis ANOVAs to assess the effect of similarity on transfer for each problem type. Solution times in the experimental conditions did not differ significantly for transfer problems of Type A, $H(3, N = 78) = 1.91, p = .59$, and Type B, $H(3, N = 78) = 5.71, p = .13$. However, transfer problems of Type C were solved faster when they were more similar to the problem solved in Block 1, $H(3, N = 78) = 11.05, p < .05$. A similar effect was seen in problems of Type D, but it was only marginally significant, $H(3, N = 78) = 6.88, p = .08$. So, there was some effect of similarity on transfer in the problem types that were most difficult in Block 1. Note, however, that problems of Types C and D were solved very fast in the transfer block. Even in the conditions of lowest similarity, the solution times for those problem types were faster than the solution times for Problem Types A and B in the conditions of highest similarity.

To test our transfer predictions, we compared solution times in the first and second blocks for each problem type. As Tables 8 and 9 show, solution times were shorter in Block 2 for all problem types. Problems with longer solution times in Block 1 showed larger transfer effects. Single Wilcoxon tests showed significant transfer for problems of Type A, $N = 78, T = 1{,}108, p < .05$; Type B, $N = 78, T = 699, p < .001$; Type C, $N = 78, T = 81, p < .001$; and Type D, $N = 78, T = 95, p < .001$.

Discussion

The results replicated the main findings from Experiments 1A and 1B. Problem type, as defined by the need to relax constraints and decompose chunks, was the main determinant of problem difficulty in the first block. The predicted order of problem difficulty was present in the analysis of solution frequency as well as in the analysis of solution time. As in Experiment 1B, there was one nonsignificant outcome; that is, problems of Type B were not solved significantly less frequently than problems of Type A within the given time limit. At the same time, there was a statistically reliable difference in solution time between these two problem types. Therefore, it is likely that the nonsignificant outcome is due to a ceiling effect (96% of the Type A and 90% of the Type B problems were solved).

The transfer effects between the first and the second block of problems followed the same pattern as that observed in Experiment 1A. Similarity had no measurable effect on solution frequency, whereas the size of the transfer effect varied as a function of initial problem difficulty. Problem types that were more difficult when encountered for the first time yielded larger transfer effects. The only nonsignificant outcome was the lack of the predicted difference in solution frequency between a first and a second encounter of problems of Type B. We think that this is also due to a ceiling effect. All in all, the analysis of solution frequency suggests that the transfer effects are mainly due to representational change.

The pattern of solution times followed the pattern predicted by the representational change hypothesis. Solution times were reduced on a second encounter for all problem types. Furthermore, the reduction was larger for problem types that were more difficult on a first encounter. Unexpectedly, problems of Types C and D yielded even shorter solution times than the other problem types in Block 2. Although our theory predicted larger transfer effects for more difficult problems, it predicted a leveling of problem difficulty, not a reversal. Also, similarity between a problem solved in the first block and the corresponding transfer problem in Block 2 had a weak effect on transfer (measured by solution times) for those problem types. We offer the following ad hoc explanations for these effects: First, all numerals were the same in Type D problems. It is plausible that this distinctive feature helped speed up recognition of Type D problems. Second, the difference between the left and right sides of the equation was always equal to five in problems of Type E, and it is possible that participants noticed this.

To summarize, the results of Experiment 2 support the conclusion that constraint relaxation and chunk decomposition are stronger determinants of transfer effects in matchstick arithmetic than surface similarity between problems. Systematic variations of similarity had only weak effects. Hence, the transfer observed in matchstick arithmetic problems is mainly a function of representational change and cannot be explained by the identical elements theory alone. The replications of the main effects of Experiments 1A and 1B show that these effects generalize across levels of familiarity with Roman numerals.

Experiment 3

The purpose of Experiment 3 was to provide additional support for the hypothesis that we have identified two separate dimensions of difficulty in insight problem solving. Experiments 1A and 1B, and 2 showed that matchstick arithmetic problems of Types C and D are more difficult than problems of Types A and B (see Figures 2, 3, 5, 6, 7, and 8). According to our theory, this quantitative similarity hides a qualitative difference.

The difficulty of Type C problems is due to the need to relax constraints on meaningful arithmetic statements, so that equations of the familiar regular form

$$X = Y + Z$$

can be transformed into equations of the unfamiliar tautological form

$$X = Y = Z.$$

The difficulty of Type D problems, in contrast, is caused by the need to decompose perceptual chunks like V and X into their components so that the problem solver can consider changing a V symbol into an X (or vice versa) by horizontally sliding either the \ or the / component.

If problems of Types C and D are difficult for qualitatively different reasons, it should be possible to manipulate one source of difficulty without affecting the other. The fact that the transformation between the regular form and the tautological form of Type C problems is reversible provides an opportunity to test this hypothesis. Consider the problem

$$\text{VI} = \text{III} = \text{III},$$

with solution VI = III + III. The move required to solve this problem is the reverse of the move required to solve the regular Type C problems: Lift one of the matchsticks that make up the second equal sign, rotate it, and put it down again, thereby transforming the equal sign into a plus sign and, indirectly, the tautological form into the regular form.

If the tautological form constitutes the initial state of a Type C problem, constraints from arithmetic or algebra should not interfere with its solution. There are no constraints that rule out the regular form, so moves that produce the regular form should not be difficult to think of. Hence, tautological Type C problems should be easier than regular Type C problems on a first encounter. Once the problem solver has seen the tautological form, and thus implicitly been told that this form is valid and can be useful, the tautology constraint ought to be relaxed. Once relaxed, it should no longer interfere with the solving of regular Type C problems. Hence, we predicted a facilitating effect of prior exposure to tautological Type C problems on the solutions to regular Type C problems.

Our theory makes different predictions for problems of Type D. For these problems, equation structure is irrelevant. The difficulty of Type D problems is due to the need to decompose chunks like V and X into their / and \ components. Hence, tautological Type D problems should be as difficult as regular Type D problems, and prior exposure to the tautological form should have little effect on the performance of regular Type D problems. These predictions can be tested by presenting Type D problems of the following sort:

$$\text{XI} = \text{VI} = \text{VI},$$

with solution VI = VI = VI.

The same logic applies to Type A problems. Because equation structure is not the determinant of the difficulty of these problems, we predicted that the difference between the regular and tautological forms should have little impact on the performance of Type A problems. This prediction can be tested with problems like

$$\text{VI} = \text{VII} = \text{V},$$

with solution VI = VI = VI. The corresponding predictions for problems of Type B cannot be tested because there is no tautological form of this problem type.

Method

Participants. Thirty-four undergraduates (18 women and 16 men) from the University of Hamburg and the University of Munich participated in this study for course credit. They ranged in age from 19 to 42 years.

Materials, design, and procedure. There were two sets of problems. The regular set consisted of six regular problems: two of Type A, two of Type B, one of Type C, and one of Type D. The tautological set consisted of four tautological problems: two of Type A, one of Type C, and one of Type D.

The participants were randomly assigned to either of the two experimental groups. The regular-first group had 18 participants, and the tautology-first group had 16 participants. Each group solved two blocks of problems. In the regular-first group, the participants solved the problems with the regular structure in the first block and the problems with the tautological structure in the second block. In the tautology-first group, the participants solved the two sets of problems in the opposite order. Problems were randomized within each block for both groups. The experimental procedure was in other respects the same as that used in the previous three experiments.

Results

Solution frequencies. An alpha level of .05 was used for all statistical tests. Only the data from Problem Types A, C, and D were relevant for the hypotheses being tested.

Table 10 shows the solution frequencies. As observed in Experiments 1A and 2, the regular form of Type C problems was difficult for the participants. Only 61% of Type C problems were solved in Block 1 in the regular-first group. In contrast, the tautological form of these problems was easy. All tautological Type C problems were solved in Block

Table 10
Percentage Solutions in Experiment 3 by Condition, Problem Type, and Block

Condition	Block 1	Block 2
Type A		
Regular-first	94	100
Tautology-first	97	91
Type C		
Regular-first	61	100
Tautology-first	100	94
Type D		
Regular-first	61	94
Tautology-first	69	93

Table 11
Solution Times (in Seconds) in Experiment 3 for Regular Problems

Group and statistic	Problem type A	C	D
Regular-first			
Mdn	32	175	193
Lower *Q*	20	113	83
Upper *Q*	59	300	300
Tautology-first			
Mdn	37	11	46
Lower *Q*	17	7	21
Upper *Q*	79	17	135

1 in the tautology-first group. The difference between the groups in Block 1 was significant $\chi^2(1, N = 34) = 12.59$, $p < .001$. As predicted by the constraint relaxation hypothesis, tautological Type C problems were easier than regular Type C problems on a first encounter.

There was transfer from the tautological to the regular Type C problems. All Type C problems were solved in Block 2 in the regular-first group. The difference between the two blocks was significant for that group, $\chi^2(1, N = 18) = 13.85$, $p < .001$. This result is in accordance with our hypothesis that prior exposure to tautological Type C problems removed the constraint that hinders the solution of regular Type C problems.

The impact of prior exposure to tautological problems on the solutions to regular problems could not be ascertained because of a ceiling effect. All problems were solved in Block 1 for that group, leaving no room for improvement on Block 2.

In contrast, the initial difficulty of Type D problems was almost unaffected by the difference between regular and tautological equation forms. In Block 1, 61% of the problems were solved in the regular-first group and 69% in the tautology-first group, a nonsignificant difference, $\chi^2(1, N = 34) = 0.22$, $p = .64$. Performance improved in both groups by approximately the same amount. In Block 2, 92% versus 94% of the problems were solved. The effect was significant for the regular-first group, $\chi^2(1, N = 18) = 5.79$, $p < .05$, and marginally significant for the tautology-first group, $\chi^2(1, N = 16) = 3.28$, $p = .07$. These results are consistent with the hypothesis that the tautological form does not affect initial difficulty and transfer in problems of Type D.

As predicted, the initial difficulty of Type A problems was also unaffected by the experimental manipulation. The solution percentages for Block 1 were 94% and 97% in the two groups. Because of the ceiling effect, the impact of prior exposure to tautological problems on the solution rate for regular problems could not be ascertained.

Solution times. As in the former experiments, we report medians and quartiles and used ordinal tests in our analysis because the distribution of solution times was far from normal. The median solution times and the lower and upper quartiles for the regular problems in both groups are shown in Table 11. The regular-first group once again replicated the predicted ordering of the three relevant Problem Types A, C, and D. Two Wilcoxon tests showed that Type C problems were solved significantly slower than Type A problems, $n = 18$, $T = 3$, $p < .001$. Moreover, Type D problems were solved significantly slower than Type A problems, $n = 18$, $T = 17, p < .01$.

The pattern of results in the tautology-first group was quite different. Solution times for the regular Type A and Type D problems were longer than those for Type C problems. Two Wilcoxon tests showed that the difference between Type A and Type C problems was significant, $n = 16$, $T = 16, p < .01$, as was the difference between Type C and Type D problems, $n = 16$, $T = 16$, $p < .01$. The difference between Type A and Type D problems was not significant ($p = .50$). Prior exposure to tautological problems made regular Type C problems easier than either Type A or Type D problems.

To further examine the impact of prior exposure to the tautological form on the solution times for regular problems, we compared the solution times from Block 1 in the regular-first condition with the solution times from Block 2 in the tautology-first condition. Single U tests showed that the difference between these conditions was not significant for Type A problems, $U = 128$, $Z = -0.53$, $p = .59$. However, the difference was significant for Type C problems, $U = 8, Z = 4.67, p < .001$, and Type D problems, $U = 65$, $Z = 2.71$, $p < .01$. Prior exposure to tautological problems caused a facilitating effect for Type C problems and Type D problems but not for Type A problems.

The median solution times and the lower and upper quartiles for the tautological problems in both experimental groups are shown in Table 12. As predicted, tautological Type C problems were easier than regular Type C problems on a first exposure, $U = 28, Z = 4.02, p < .001$. Tautological Type D problems were as difficult as regular Type D problems on an initial encounter; the numerical difference was far from significant, $U = 116$, $Z = 0.96$, $p = .33$. Unexpectedly, participants took somewhat longer to solve tautological Type A problems than regular Type A problems; the difference was marginally significant, $U = 34$, $Z = -1.17, p = .09$.

We conducted further Wilcoxon tests to determine if there were significant differences between the three problem types in the tautology-first condition. Tautological problems of

Table 12
Solution Times (in Seconds) in Experiment 3 for Tautological Problems

Group and statistic	Problem type A	C	D
Tautology-first			
Mdn	79	27	106
Lower *Q*	23	17	33
Upper *Q*	111	53	300
Regular-first			
Mdn	12	18	21
Lower *Q*	9	13	12
Upper *Q*	26	35	110

Type C were solved significantly faster than problems of Type A, $n = 16, T = 27, p < .05$, and problems of Type D, $n = 16, T = 27, p < .05$. Although problems of Type A were solved faster than problems of Type D, the difference did not reach statistical significance, $n = 16, T = 38, p = .21$.

To examine the impact of prior exposure to the regular form on the solution times for tautological problems, we compared the solution times from Block 1 in the tautology-first condition with the solution times from Block 2 in the regular-first condition. Tautological Type A problems were solved significantly faster when regular Type A problems had been encountered earlier, $U = 54, Z = -3.10, p < .01$. Tautological Type D problems were solved marginally significantly faster when regular Type D problems had been encountered earlier, $U = 89, Z = -1.88, p = .06$. Solving regular Type C problems first had no impact on the solution times of later encountered tautological Type C problems, $U = 113, Z = -1.07, p = .28$.

Discussion

The results of Experiment 3 bore out the predictions that the initial difficulty of Type C problems would be affected by the tautological form, whereas the initial difficulty of tautological Type D problems would be unaffected, and that prior exposure to tautological problems would have a different effect on regular problems of Type C than on problems of either Type A or Type D.

The solution frequencies and the solution times support the prediction that tautological Type C problems were less difficult than regular Type C problems on a first encounter. They were even easier than the tautological form of Type A and Type D problems. In contrast, tautological problems of Type D were as difficult as regular Type D problems. Tautological problems of Type A were solved as often as regular Type A problems, but unexpectedly, the solution times for tautological problems were significantly higher than the solution times for regular problems. One possible ad hoc explanation for this effect is that problem solvers first tried to manipulate the equal signs in the tautological form because their initial goal included the constraint that the solution consisted of the regular form. Only if the manipulation of the equal signs did not lead to a solution did they start to manipulate values within the irregular form.

The results also support our predictions about the effects of prior exposure to tautological problems on the solution of regular problems. Regular problems of Type C were solved more often and faster if tautological Type C problems had been solved before, whereas the difficulty of tautological Type C problems was unaffected by prior solution of regular Type C problems. There was transfer in both directions between the tautological and the regular Type D problems (with the small restriction that the transfer was only marginally significant in the tautology-first group). Although solving tautological Type A problems first had no impact on the later solution of regular Type A problems, solving regular Type A problems first helped to solve tautological Type A problems later. The latter result was unexpected, but it can be understood by the same logic that explains the higher difficulty of the tautological Type A problems. Problem solvers who have solved the regular form of Type C problems know that it can be useful to produce the tautological form. Therefore, they do not prefer moves that produce the regular form, whereas those problem solvers who do not have experience with the regular form do.

General Discussion

The constraint relaxation and chunk decomposition hypotheses are powerful tools for understanding both the relative difficulty of individual problems and transfer effects. Table 13 gives an overview of the effects of task difficulty observed in Experiments 1A, 1B, and 2.

Our analysis of the need to relax constraints predicted the rank ordering of three types of superficially similar matchstick arithmetic problems with respect to initial difficulty. We predicted that problems of Type A would be less difficult than problems of Type B, which in turn would be less difficult than problems of Type C (see Tables 1, 4, and 6 for several instances of each problem type). Rows 1, 2, 4, and 5 in Table 13 illustrate that the predicted order was observed in all three experiments that assessed the main effect of task difficulty (Experiments 1A, 1B, and 2) in both dependent variables—solution frequency and solution time. Problems of Types A and B did not differ significantly in the analysis of solution frequencies in Experiments 1B and 2. However, at the same time, the differences in solution times were highly significant. Therefore, it is likely that the nonsignifi-

Table 13
Main Effects of Task Difficulty in Experiments 1A, 1B, and 2

	Experiment 1A		Experiment 1B		Experiment 2	
Prediction	Predicted order	p	Predicted order	p	Predicted order	p
% solved						
A > B	Yes	<.05	Yes	.12	Yes	.14
B > C	Yes	<.01	Yes	<.001	Yes	<.001
A > D	Yes	.13	Yes	<.05	Yes	<.001
Time						
A < B	Yes	<.001	Yes	<.001	Yes	<.01
B < C	Yes	<.05	Yes	<.01	Yes	<.001
A < D	Yes	<.001	Yes	<.001	Yes	<.001

cant outcomes in the analysis of solution frequency are due to ceiling effects (more than 75% of the Type B problems were solved in Experiments 1B and 2).

The chunk decomposition hypothesis predicts that problems that require the decomposition of tight perceptual chunks (i.e., chunks with components that are not themselves chunks) are more difficult than problems that require the decomposition of intermediate and loose chunks (i.e., chunks with components that are themselves chunks). This led to the prediction that problems of Type D would be more difficult than problems of Type A (see Tables 1, 4, and 6 for several instances of each problem type). Rows 3 and 6 in Table 13 show that the predicted order of task difficulty was present in Experiments 1A, 1B, and 2, in both solution frequencies and solution times. The solution frequencies in Experiment 1A did not differ significantly between problems of Types D and A. Again, the difference in solution time was highly significant, making it likely that the nonsignificant outcome is due to a ceiling effect and the relatively small sample size in Experiment 1.

The constraint relaxation and chunk decomposition hypotheses also predict large transfer effects for the more difficult problem types, namely, Types B, C, and D. Table 14 gives an overview of the transfer effects obtained in Experiments 1A and 2.

The predicted order of task difficulty for the first and second encounters of a problem type was never violated. Although transfer for the easiest Type A problems was seen only in the solution times of Experiment 2 (probably because of the large sample size), significant transfer for problems of Types B, C, and D was present in Experiments 1A and 2 in both solution frequencies and solution times. There were two deviations from this general pattern. There was no significant effect of transfer on solution frequencies for problems of Type D in Experiment 1 and problems of Type B in Experiment 2. However, we observed significant differences in solution time in both cases. The lack of significant effects is once again likely to be due to ceiling effects. The solution frequency for Type D problems in Experiment 1 was already 75% during the first block, and the sample size was quite small. The solution frequency for Type B problems in Experiment 2 was more than 80%. Therefore, there was little room for improvement in the second block.

The constraint relaxation and chunk decomposition hypotheses predict not only transfer for more difficult problem types but also an interaction such that problems with higher initial difficulty exhibit larger transfer effects. We found this interaction in all four experiments. The effect of task difficulty observed on a first encounter always disappeared when the same problem types were encountered a second time. The leveling out of task difficulty on a second encounter is best seen when comparing Figures 2 and 3 with Figure 4 (Experiment 1A) and comparing Figures 7 and 8 with Figure 9 (Experiment 2).

Experiment 3 provided additional support for the hypothesis that the need to relax constraints and the need to decompose chunks are different sources of difficulty in insight problem solving. Although presenting problems of a tautological form completely removed the former source of difficulty, it did not affect the latter.

To summarize, our experiments provide some evidence that constraint relaxation and chunk decomposition are the main determinants of task difficulty and transfer in the domain of matchstick arithmetic problems. The predicted order of task difficulty was never violated, and the predicted transfer was always present, at least numerically. The few nonsignificant results we obtained are likely due to ceiling effects. In the next section, we discuss whether alternative theories of insight can explain our results.

Table 14
Transfer in Experiments 1A and 2

Prediction	Experiment 1A: Predicted order	Experiment 1A: p	Experiment 2: Predicted order	Experiment 2: p
% solved				
A (1) ≤ A (2)	Yes	.80	Yes	.89
B (1) < B (2)	Yes	<.05	Yes	.14
C (1) < C (2)	Yes	<.001	Yes	<.001
D (1) < D (2)	Yes	.13	Yes	<.001
Time				
A (1) ≥ A (2)	Yes	.91	Yes	<.05
B (1) > B (2)	Yes	<.001	Yes	<.001
C (1) > C (2)	Yes	<.01	Yes	<.001
D (1) > D (2)	Yes	<.05	Yes	<.001

Note. Numbers in parentheses refer to experiment numbers (1 = Experiment 1A; 2 = Experiment 2).

Alternative Explanations

Theories of insight fall into two broad classes: theories that explain why insight is at all possible and theories that explain why insight is difficult to attain. Also, there are theories of thinking based on alternative assumptions about knowledge representation.

How is insight possible? According to the Gestalt psychologists (Köhler, 1924, 1925; Wertheimer, 1945/1959), a solution appears in consciousness when the perceptual field reorganizes itself into a better, more harmonious or balanced state (Ohlsson, 1984a). In the terminology of contemporary complexity theory, the Gestalt psychologists (especially Köhler, 1924) saw insight as an instance of self-organization.

At a general level, the gestalt concept of reorganization is certainly applicable to matchstick arithmetic. However, it cannot explain particular effects. For example, why does reorganization of the perceptual field take longer in problems of Types C and D than in problems of Type A? The reorganization concept has not been specified at the level of detail that is required to answer such questions.

The hypothesis of variation and selection claims that novel solutions are constructed by creating more or less random variants of existing solutions until one is found that solves the current problem (Johnson-Laird, 1989). This idea has been proposed many times under different names: blind variation and selective retention (Campbell, 1960), chance permutations and configuration formation (Simonton, 1988), generate and test (Newell & Simon, 1972), generation and exploration (Finke, 1995; Finke, Ward, & Smith, 1992), and trial and error (Thorndike, 1903, 1913).

As a general characterization of problem solving, the variation–selection hypothesis is certainly correct. What else can a problem solver do if one solution fails than try another one? In particular, it is plausible that our participants kept generating different possible moves until they found one that worked. However, without auxiliary assumptions that predict the order in which the possible moves are considered, the variation–selection principle cannot explain why it takes longer to generate the right move for some problems than others. Also, it is not clear how the variation–selection idea by itself can explain differential transfer effects for different types of problems.

In contrast to both the reorganization and variation–selection hypotheses, Bowers, Farfolden, and Mermigis (1995), Perkins (1981), Weisberg (1986, 1993, 1995), Weisberg and Alba (1981), and others have described creative problem solving as a gradual transformation of past experience. Detailed analyses of creative processes have revealed how artistic and creative products are rooted in such experience (Weisberg, 1993).

The gradual transformation view is implausible with respect to matchstick arithmetic. Our participants solved the experimental problems in a matter of seconds or minutes, and, unlike the artists and scientists studied by Weisberg (1993) and others, they did not produce and repeatedly revise a creative product. The gradual transformation hypothesis is consistent with what is known about temporally extended creative enterprises, but it does not explain why some single-step problems take longer to solve than others nor why such problems exhibit differential transfer effects.

These three hypotheses fail to explain our data because they do not provide any mechanism for explaining why impasses occur and hence cannot explain why an impasse lasts longer in one problem than in another. We next consider theories explaining impasses but not insights.

Why is insight difficult? The functional fixedness hypothesis claims that mental representations of objects are associated with the common functions for those objects. An insight problem that requires a familiar object to be used in an unfamiliar way is difficult to solve because the unfamiliar usage of the object is blocked by the spontaneous retrieval of its more familiar function. This hypothesis has empirical support (Adamson, 1952; Birch & Rabinowitz, 1951; Duncker, 1945; Flavell, Cooper, & Loiselle, 1958). Two computer models of functional fixedness have been described in Keane (1989) and in an unpublished report by Greeno and Berger (1987); both are reviewed in Ohlsson (1992b, pp. 25–27).

With respect to matchstick arithmetic, the functional fixedness hypothesis suggests that our participants did not immediately think of the right move because they were distracted by spontaneous recall of more familiar uses of matches or spontaneous recall of the normal meaning of arithmetic symbols. This hypothesis is plausible but cannot explain the differences among problem types with respect to either initial difficulty or amount of transfer because all of the problems involved matches. Also, the functional fixedness hypothesis explains why impasses arise but not how they are resolved. People must have some process by which they eventually come to think of the unfamiliar function; otherwise they would fail to solve insight problems in general and matchstick arithmetic problems in particular, contrary to our observations.

The mental ruts hypothesis claims that insight problems are difficult because initial exploration of an unsuccessful search path adds activation to that path, which in turn increases the probability of further exploration of that path and therefore decreases the probability of thinking of another path. This type of explanation has been expressed in different terminologies: Einstellung (Luchins, 1942), familiarization and selective forgetting (Simon, 1966), mental ruts (Smith, 1995a, 1995b), set and forgetting (Weisberg, 1986, p. 30), and false starts and absence of interference (Woodworth, 1938, p. 823).

The mental ruts hypothesis claims that our participants did not immediately think of the right move because they kept considering the same wrong move (or moves) over and over again. This is indeed a potential cause of impasses. However, the mental ruts hypothesis also predicts that problem solvers, once entrenched in a false solution, cannot recover without a pause. The main way to escape from a mental rut is to set the problem aside for a while so that the activation of the unsuccessful path can decay to the point where alternative paths can compete during retrieval and decision making (Ohlsson, 1992b). But our participants did not have the opportunity to take a pause; yet, they often did resolve their impasses. Another possibility is that an external stimulus shakes the problem-solving process out of a mental rut. But the display never changed while participants solved a task, so the information in the problem environment remained constant. Also, the mental ruts hypothesis cannot explain why the tendency to become stuck in a mental rut is stronger for some matchstick arithmetic problems than others. Nor can this hypothesis explain differential transfer effects for different problem types.

Alternative knowledge representations. Other alternative hypotheses differ from ours by assuming that the relevant knowledge is encoded in some other format than constraints and chunks. For example, the remote associations hypothesis claims that the participants have an associative hierarchy for matchstick arithmetic problems such that certain actions or operations are more likely to be accessed than others (Smith, 1995a, 1995b). Priming of the relevant associations might produce short-term transfer effects.

It is highly unlikely that the majority of the participants in our experiments had a well-developed hierarchy of associations for matchstick arithmetic problems. Although some of them might conceivably have seen a matchstick problem in a psychology textbook or a collection of brain teasers, solving matchstick problems is not, in our experience, a common pastime among college students either in the United States or in Germany. Also, an explanation for our data in terms of association hierarchies requires auxiliary assumptions to explain why the associations needed to solve some problem types are more remote than others.

The related differential recall hypothesis claims that task knowledge is encoded in procedures (e.g., collections of production rules; Anderson & Lebiere, 1998) and that different procedures are associated with different speeds because of different degrees of familiarity. Although the participants in our experiments probably did not possess specific procedures for matchstick arithmetic problems, they obviously did possess general procedures or motor programs for manipulating physical objects. To explain our findings, this hypothesis must be augmented with auxiliary assumptions to explain why procedures for sliding or rotating an object are recalled slower than procedures for moving an object. Everyday experience contains enough instances of objects being rotated or pushed sideways so that this hypothesis is an unlikely explanation for why one set of procedures would take tens of seconds—in some cases several minutes—longer to recall.

Generality

The hypotheses that impasses are caused by unhelpful constraints and chunks acquired in past experience and that impasses are resolved through constraint relaxation and chunk decomposition presuppose that at least some knowledge is encoded in constraints and chunks. Our theory would be strengthened by independent evidence for this assumption. At the task-specific level, we are certain that the participants in our experiments initially assimilated matchstick arithmetic problems to their prior knowledge of arithmetic; the number symbols, the arithmetic operations, and the equation format make this highly plausible. However, we do not have independent evidence that this knowledge is encoded in constraints. Our chunks hypotheses are on stronger ground. Our training procedure (see the *Method* section in Experiment 1A) explicitly instructed the participants to see symbols like *VII* as complex patterns and *I* as a subpattern of *VII*, and it is highly likely that their past experience had biased them against seeing / and \ as meaningful subpatterns of *V* and *X*. In short, our experiments tested the specific form that our general hypotheses took when articulated vis-à-vis the domain of matchstick arithmetic problems, as opposed to the general hypotheses themselves.

However, the generality of constraints and chunks is supported by a wide range of studies. Constraints that need to be relaxed for successful problem solving have been identified in many other insight problems than matchstick arithmetic. For example, the Nine Dots Problem[1] is difficult because people try to keep their lines inside the square formed by the dots, a constraint that prohibits the solution from being found (Scheerer, 1963, p. 120).[2] The Six Matches Problem[3] is difficult because people implicitly assume that the figure is supposed to be two-dimensional (Bogoyavlenskaya, 1972; Scheerer, 1963, p. 120). Similarly, it is plausible that the Inverted Pyramid Problem[4] is difficult because people implicitly assume that the $100 bill is not to be destroyed. The source of this constraint in everyday life is obvious. Isaak and Just (1995) specified the constraints that need to be relaxed for 21 different insight problems (Table 9.1, p. 284). The constraint construct is also useful in explaining how people solve the Tower of Hanoi puzzle[5] (Richard et al., 1993), how problem solvers identify the most fruitful mapping between a target and many possible sources in analogical reasoning (Holyoak & Thagard, 1989), how learners detect and correct their own errors during skill acquisition (Mitrovich & Ohlsson, in press; Ohlsson, 1992a, 1996; Ohlsson, Ernst, & Rees, 1992; Ohlsson & Rees, 1991), and how cognitive dissonance is reduced (Shultz & Lepper, 1996).

The usefulness of the chunk construct for understanding noninsight problem solving was demonstrated by Chase and Simon (1973). They found that chess masters perceive a complex chessboard in terms of well-known configurations of chess pieces rather than in terms of single chessmen, as the novice player does. These results have been replicated with respect to other board games (Reitman, 1976), electronic circuit diagrams (Egan & Schwartz, 1979), and other domains (Ericsson, 1985). Chunking effects also have been observed in geometry proof finding (Koedinger & Anderson, 1990; Ohlsson, 1990). The chunk construct is also a central component in explanations of skill acquisition (Rosenbloom et al., 1989), training effects in working memory (see Ericsson & Lehmann, 1996, pp. 292–296, for a review), the perception of reversible figures (Attneave, 1971; Hochberg & Brooks, 1960; Toppino & Long, 1987), and social stereotypes (Hamilton, 1994; Hilton & Hippel, 1996).

In short, the two constructs of constraints and chunks are useful in understanding a wide variety of psychological phenomena. An explanation of insight in terms of constraint relaxation and chunk decomposition is thus consistent with a large body of psychological research.

In general, the concept of representational change is consistent with the fact that every problem is both like and unlike past problems. Hence, human beings have no choice but to extrapolate the past in their efforts to make sense of the present, and in most situations, past experience is indeed more help than hindrance. However, there is no guarantee that any particular extrapolation will succeed, and the complexity and the variability of the environment ensure that individuals will encounter problems and situations in which past experience is misleading. To act successfully vis-à-vis such situations, the mind must override the imperatives of experience by changing its representation of the

[1] Nine dots form a 3×3 square on a piece of paper. Draw a line with a pencil that goes through all nine dots, without duplicating any part of the path and without lifting the pencil from the paper.

[2] This interpretation of performance on the Nine Dots Problem has been criticized by Weisberg and Alba (1981, 1982). We disagree with their critique for reasons stated in Ohlsson (1992b, p. 15).

[3] Construct four equilateral triangles out of six matches.

[4] A steel pyramid is perfectly balanced upside down on a steel table, its tip resting on a $100 bill. Remove the bill without disturbing the pyramid.

[5] There are N disks stacked in order of decreasing size on one of three pegs. Move the stack to another peg by moving one disk at a time and without ever putting a larger disk on a smaller one.

problem by relaxing constraints; decomposing chunks; and, presumably, by other, yet to be discovered, processes.

References

Adamson, R. E. (1952). Functional fixedness as related to problem solving. *Journal of Experimental Psychology, 44,* 288–291.

Anderson, J. R. (1989). The analogical origins of errors in problem solving. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing* (pp. 343–371). Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Attneave, F. (1971). Multistability in perception. *Scientific American, 225,* 62–71.

Birch, H. G., & Rabinowitz, H. S. (1951). The negative effect of previous experience on productive thinking. *Journal of Experimental Psychology, 41,* 121–125.

Bogoyavlenskaya, D. B. (1972). Replacement of the model of the problem situation in the problem-solving process. In V. N. Pushkin (Ed.), *Problems of heuristics* (pp. 114–171). Jerusalem: Israel Program for Scientific Translations.

Bowers, K. S., Farfolden, P., & Mermigis, L. (1995). Intuitive antecedents of insight. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 73–95). Cambridge, MA: MIT Press.

Bowers, K. S., Regehr, G., Balthazard, C., & Parker, K. (1990). Intuition in the context of discovery. *Cognitive Psychology, 22,* 72–110.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer, Jr. (Eds.), *Advances in social cognition* (Vol. 1, pp. 1–36). Hillsdale, NJ: Erlbaum.

Campbell, D. T. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review, 67,* 380–400.

Chase, W., & Simon, H. A. (1973). The mind's eye in chess. In W. Chase (Ed.), *Visual information processing* (pp. 251–281). New York: Academic Press.

Chi, M. T. H. (1992). Conceptual change within and across ontological categories: Examples from learning and discovery in science. In R. N. Giere (Ed.), *Cognitive models of science* (pp. 129–186). Minneapolis: University of Minnesota Press.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–18.

Dorfman, J., Shames, V. A., & Kihlstrom, J. F. (1996). Intuition, incubation, and insight: Implicit cognition in problem solving. In G. Underwood (Ed.), *Implicit cognition* (pp. 257–296). Oxford, England: Oxford University Press.

Duncker, K. (1945). On problem solving. *Psychological Monographs, 68*(5, Whole No. 270).

Egan, D., & Schwartz, B. (1979). Chunking in the recall of symbolic drawings. *Memory & Cognition, 7,* 149–158.

Ericsson, K. A. (1985). Memory skill. *Canadian Journal of Psychology, 39,* 188–231.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47,* 273–305.

Finke, R. A. (1995). Creative insight and preinventive forms. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 255–280). Cambridge, MA: MIT Press.

Finke, R., Ward, T., & Smith, S. (1992). *Creative cognition.* Cambridge, MA: MIT Press.

Fiore, S. M., & Schooler, J. W. (1998). Right hemisphere contributions to creative problem solving: Converging evidence for divergent thinking. In M. Beeman & C. Chiarello (Eds.), *Right hemisphere language comprehension* (pp. 349–371). Mahwah, NJ: Erlbaum.

Flavell, J. H., Cooper, A., & Loiselle, R. H. (1958). Effect of number of pre-utilization functions on functional fixedness in problem solving. *Psychological Reports, 4,* 343–350.

Greeno, J. G., & Berger, D. (1987). *A model of functional knowledge and insight* (Tech. Rep. No. GK-1). Berkeley: University of California, School of Education.

Hamilton, D. (1994). Stereotypes. In R. Wyer & T. Srull (Eds.), *Handbook of social cognition: Applications* (2nd ed., Vol. 2, pp. 1–68). Hillsdale, NJ: Erlbaum.

Helmholz, H. von. (1896). *Vorträge und reden* [Lectures and talks, Vol. 1]. Brunswick, Germany: Friedrich Viehweg und Sohn.

Hilton, J., & Hippel, W. von. (1996). Stereotypes. *Annual Review of Psychology, 47,* 237–271.

Hochberg, J., & Brooks, V. (1960). The psychophysics of form: Reversible-perspective drawings of spatial objects. *American Journal of Psychology, 73,* 337–354.

Holding, D. H. (1985). *The psychology of chess skill.* Hillsdale, NJ: Erlbaum.

Holyoak, K., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13,* 295–355.

Isaak, M. I., & Just, M. A. (1995). Constraints on thinking in insight and invention. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 281–325). Cambridge, MA: MIT Press.

Johnson-Laird, P. N. (1989). Analogy and the exercise of creativity. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 313–331). New York: Cambridge University Press.

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

Keane, M. (1989). Modelling problem solving in Gestalt "insight" problems. *Irish Journal of Psychology, 10,* 201–215.

Keane, M. (1996). On adaptation in analogy: Tests of pragmatic importance and adaptability in analogical problem solving. *Quarterly Journal of Experimental Psychology, 49A(4),* 1062–1085.

Knoblich, G., & Haider, H. (1996). Empirical evidence for constraint relaxation in insight problem solving. In G. W. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 580–585). Mahwah, NJ: Erlbaum.

Knoblich, G., & Wartenberg, F. (1998). Unbemerkte Lösungshinweise begünstigen Veränderungen der Problemrepräsentation [Unnoticed hints facilitate representational change in problem solving]. *Zeitschrift für Psychologie, 206,* 207–234.

Koedinger, K., & Anderson, J. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science, 14,* 511–550.

Koffka, K. (1935). *Principles of Gestalt psychology.* London: Routledge & Kegan Paul.

Köhler, W. (1924). *Die physischen gestalten in ruhe und im stationären zustand* [The physical shapes in rest and in stationary state]. Erlangen, Germany: Verlag der Philosophischen Akademie.

Köhler, W. (1925). *The mentality of apes.* New York: Liveright.

Luchins, A. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs, 54*(Whole No. 6).

Lung, C. T., & Dominowski, R. L. (1985). Effect of strategy instructions and practice on nine-dot problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 804–811.

Metcalfe, J. (1986a). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 288–294.

Metcalfe, J. (1986b). Premonitions of insight predict impending error. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 623–634.

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

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Mitrovich, A., & Ohlsson, S. (in press). Evaluation of a constraint-based tutor for a database language. *International Journal of Artificial Intelligence and Education.*

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.

Ohlsson, S. (1984a). Restructuring revisited: I. A summary and critique of the Gestalt theory of problem solving. *Scandinavian Journal of Psychology, 25,* 67–78.

Ohlsson, S. (1984b). Restructuring revisited: II. An information processing theory of restructuring and insight. *Scandinavian Journal of Psychology, 25,* 117–129.

Ohlsson, S. (1990). The mechanism of restructuring in geometry. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 237–244). Hillsdale, NJ: Erlbaum.

Ohlsson, S. (1992a). Constraint-based student modeling. *Journal of Artificial Intelligence and Education, 3,* 429–447.

Ohlsson, S. (1992b). Information-processing explanations of insight and related phenomena. In M. Keane & K. Gilhooley (Eds.), *Advances in the psychology of thinking* (pp. 1–44). London: Harvester-Wheatsheaf.

Ohlsson, S. (1996). Learning from performance errors. *Psychological Review, 103,* 241–261.

Ohlsson, S., Ernst, A. M., & Rees, E. (1992). The cognitive complexity of learning and doing arithmetic. *Journal for Research in Mathematics Education, 23,* 441–467.

Ohlsson, S., & Rees, E. (1991). The function of conceptual understanding in the learning of arithmetic procedures. *Cognition and Instruction, 8,* 103–179.

Patalano, A. J., & Seiffert, C. M. (1994). Memory for impasses during problem solving. *Memory & Cognition, 22,* 234–242.

Perkins, D. (1981). *The mind's best work.* Cambridge, MA: Harvard University Press.

Poincaré, H. (1952). *Science and method.* New York: Dover. (Original work published 1908).

Purdue, C. W., & Gurtman, M. B. (1990). Evidence for the automaticity of ageism. *Journal of Experimental Social Psychology, 26,* 199–216.

Reitman, J. (1976). Skilled perception in Go: Deducing memory structures from inter-response times. *Cognitive Psychology, 8,* 336–356.

Richard, J.-F., Poitrenaud, S., & Tijus, C. (1993). Problem-solving restructuration: Elimination of implicit constraints. *Cognitive Science, 17,* 497–529.

Rokeach, M. (1970). *Beliefs, attitudes and values.* San Francisco: Jossey-Bass.

Rosenbloom, P., Laird, J., & Newell, A. (1989). The chunking of skill and knowledge. In B. Elsendoorn & H. Bouma (Eds.), *Working models of human perception* (pp. 391–410). London: Academic Press.

Scheerer, M. (1963). Problem solving. *Scientific American, 208,* 118–128.

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

Schultz, T. R., & Lepper, M. R. (1996). Cognitive dissonance reduction as constraint satisfaction. *Psychological Review, 103,* 219–240.

Simon, H. A. (1966). Scientific discovery and the psychology of problem solving. In R. Colodny (Ed.), *Mind and cosmos* (pp. 22–40). Pittsburgh, PA: University of Pittsburgh Press.

Simonton, D. K. (1988). *Scientific genius.* Cambridge, England: Cambridge University Press.

Singley, M., & Anderson, J. (1989). *The transfer of cognitive skill.* Cambridge, MA: Harvard University Press.

Smith, S. M. (1995a). Fixation, incubation, and insight in memory and creative thinking. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 135–156). Cambridge, MA: MIT Press.

Smith, S. M. (1995b). Getting into and out of mental ruts: A theory of fixation, incubation and insight. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 229–251). Cambridge, MA: MIT Press.

Smith, S. M., Ward, T. B., & Finke, R. A. (Eds.). (1995). *The creative cognition approach.* Cambridge, MA: MIT Press.

Sternberg, R. J., & Davidson, J. E. (Eds.). (1995). *The nature of insight.* Cambridge, MA: MIT Press.

Thorndike, E. L. (1903). *Educational psychology.* New York: Lemcke and Buechner.

Thorndike, E. L. (1913). *Educational psychology* (Vol. 2). New York: Columbia University, Teachers College.

Toppino, T. C., & Long, G. M. (1987). Selective adaptation with reversible figures: Don't change that channel. *Perception & Psychophysics, 42,* 37–48.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4,* 45–69.

Wallas, G. (1926). *The art of thought.* New York: Harcourt Brace Jovanovich.

Weisberg, R. W. (1986). *Creativity, genius and other myths.* New York: Freeman.

Weisberg, R. W. (1993). *Creativity.* New York: Freeman.

Weisberg, R. W. (1995). Case studies of creative thinking: Reproduction versus restructuring in the real world. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 53–72). 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,* 169–192.

Weisberg, R. W., & Alba, J. W. (1982). Problem solving is not like perception: More on Gestalt theory. *Journal of Experimental Psychology: General, III,* 326–330.

Wertheimer, M. (1959). *Productive thinking.* New York: Harper & Row. (Original work published 1945).

Woodworth, R. (1938). *Experimental psychology.* New York: Henry Holt.

Yaniv, I., & Meyer, D. E. (1987). Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving. *Experimental Psychology: Learning, Memory, and Cognition, 13,* 187–205.

Received May 15, 1998
Revision received May 10, 1999
Accepted May 10, 1999 ■