

Expertise as mental set: The effects of domain knowledge in creative problem solving

JENNIFER WILEY

University of Pittsburgh, Pittsburgh, Pennsylvania

Experts generally solve problems in their fields more effectively than novices because their well-structured, easily activated knowledge allows for efficient search of a solution space. But what happens when a problem requires a broad search for a solution? One concern is that subjects with a large amount of domain knowledge may actually be at a disadvantage, because their knowledge may confine them to an area of the search space in which the solution does not reside. In other words, domain knowledge may act as a mental set, promoting fixation in creative problem-solving attempts. A series of three experiments in which an adapted version of Mednick's (1962) remote associates task was used demonstrates conditions under which domain knowledge may inhibit creative problem solving.

The possession of a large body of domain knowledge is central to expertise. However, the domain knowledge of experts is marked not just by its amount but also by its structure (Bédard & Chi, 1992; Ericsson & Staszewski, 1989). The organization of domain knowledge in a way that is accessible, proceduralized, integrated, and principled enables experts to excel at memory and problem-solving tasks in a number of characteristic ways. An expert can typically recognize, store, and retrieve large meaningful chunks of domain-related information. An expert's processing is also commonly more abstract or conceptual than that of a novice. The proceduralization of an expert's knowledge base tends to allow for quick and easy access to memory and possible solution paths. Further, an expert's knowledge usually contributes to better problem representation, as experts are able to engage in a more qualitative analysis of a problem—for example, by recognizing relevant features, inferring missing information, and imposing constraints in order to narrow the possible search for a solution. Over the past 20 years, these distinguishing marks of expertise have been found to be quite robust across a wide variety of domains (see Chi, Glaser, & Farr, 1988; Ericsson & J. Smith, 1991). Yet, expert performance is not uniformly superior, and the above characteristics of expert processing may sometimes lead to inferior performance in particular conditions.

When Nonexperts Outperform Experts

A number of studies have demonstrated that novices or individuals at an intermediate level of knowledge may be

seen to outperform experts. A first group of studies, related to the domain-specific nature of expertise, indicates that novices may outperform experts in conditions in which experts cannot make use of their domain knowledge. For example, chess experts' recall of randomized chess boards is slightly worse than that of novices (Chase & Simon, 1973). Similarly, Voss, Vessonder, and Spilich (1980) found that individuals with low baseball knowledge recalled a significantly greater number of baseball-irrelevant propositions from a text passage describing one half-inning of a baseball game than did individuals with high baseball knowledge. Further, studies have found no advantage for experts on tasks in which the presentation of domain-related stimuli does not correspond to typical structures of the domain, such as memory for novice-generated passages about baseball (Voss et al., 1980). Experts can also be worse at accurately judging their own performance within their domain. Glenberg and Epstein (1987) showed that physics and music experts overestimated their comprehension of a text about their area of expertise, whereas nonexperts showed a more accurate calibration. Even though the experts did comprehend the texts better than did the nonexperts, the experts were overconfident in their abilities and were less accurate than nonexperts in their estimation of their understanding of the text.

A second group of studies indicates that novices or intermediates surpass experts on tasks that require memory for the surface structure of presented information. In medical diagnosis tasks, the *intermediate effect* has been well documented (Patel & Groen, 1991; Schmidt & Boshuizen, 1993). The basic finding is that, although doctors are more accurate at diagnosis, they are worse than intermediates, such as third-year interns, at recalling or recognizing the exact information that they were presented with to make their decisions. Likewise, baseball experts are worse than novices on recognition tasks that require verbatim memory for baseball stories (Arkes & Freedman, 1984). When asked to select only the exact sentences that were actually in the text they had read, experts were more

This article is based on a dissertation submitted to the University of Pittsburgh in partial fulfillment of the requirements for a Ph.D. The author thanks her advisor, James F. Voss, and her committee, Kevin Ashley, Robert Glaser, and Jonathan Schooler, for their suggestions and support; Mara Georgi for her assistance in data coding; and Colleen Seifert, Steven Smith, and Barry Stein for their comments on an earlier version of this paper. Correspondence should be addressed to J. Wiley, Department of Psychology, University of Massachusetts, Amherst, MA 01003 (email: jwiley@psych.umass.edu).

likely than novices to choose synonymous distractors or inferences based on the presented material. Experts in computer programming (Adelson, 1984) are worse at recognizing programs they have analyzed than are novices. And, although expert radiologists have better memory for atypical (and hence diagnostic) features of X rays that they have seen, they are noticeably worse than novices at recognizing normal X rays (Myles-Worsley, Johnston, & Simons, 1988). In these cases, experts are at a disadvantage as a result of the more abstract or principled nature of their processing, which, in most domain-related circumstances, leads to better or more efficient performance.

A third group of studies indicates that experts can be outperformed by novices when a new task or context runs counter to highly proceduralized behaviors. Experts perform worse than novices when a shift from a standard means of representation is required or when a standard response is inappropriate. For example, waitresses and bartenders perform surprisingly poorly at Piaget's water-level task, as compared with groups without intensive liquid-filled container experience, presumably because the representation required to correctly solve the task is not the one that is of functional importance in their fields (Hecht & Proffitt, 1995). Instead of shifting to an appropriate way of framing the task, waitresses and bartenders continue to use the representation that is usually relevant to their job performance. Likewise, Marchant, Robinson, Anderson, and Schadewald (1991) found that expert accountants were less likely than novices to correctly apply new information they had just read about a tax law that disqualified certain standard business deductions. In another study, Frensch and Sternberg (1989) found that expert bridge players had a harder time than novices adapting to a new version of the game that changed the bidding procedure. In these studies, highly proceduralized domain knowledge, which is usually advantageous in processing domain-related information, may have led the experts to act less flexibly in new contexts.

The present study investigated another condition in which people with less domain knowledge may outperform people with a large amount of domain knowledge—namely, creative problem solving. Experts have been seen to solve problems in their fields more effectively than novices because a well-structured, easily activated knowledge base allows for efficient search of a solution space and possibly for automatic access to promising solution paths. But what happens when a problem requires a broad search for a solution outside the usual scope of the domain? One possibility is that subjects with expertise in the domain might actually be at a disadvantage, because their knowledge may confine them to an area of the search space in which the solution does not reside. In other words, domain knowledge may act as what traditionally has been called a *mental set* in the problem-solving literature, constraining search and promoting fixation in the creative problem solving of experts.

Mental Set and Fixation in Creative Problem Solving

A number of classic examples of mental set and fixation in problem solving have come from the Gestalt psychologists in their studies of *insight* problems (Duncker, 1945; Luchins, 1942; Maier, 1931). Such problems are often described as requiring productive and nonroutine thinking for a solution—that is, there must be no established procedures that could be used to guarantee a solution (as in arithmetic problems) and the solution must not be wholly formed in memory (as in the case of a problem that has been solved before). Instead, when a person approaches an insight problem, the means toward the solution are initially nonobvious, and the solution requires either a completely original approach or a novel combination of diverse bits of information through remote associations in memory (Weisberg & Alba, 1981).

In one of the first studies of mental set, Maier (1931) brought subjects into a room with two strings hanging from the ceiling more than an arm-span apart. On the floor were a chair, pliers, and a jar of tacks. The subjects were asked to tie the two strings together. One insightful solution is to use the pliers as a weight on a pendulum by tying the pliers to one string and swinging it while grasping the other string. However, this solution does not come easily to subjects who tend to see pliers as a grasping tool and not as a weight. Two factors contribute to the difficulty of solving this problem. First, the solution to this problem is impeded by the subject's mental set, which, in this case, is the tendency to consider only the normative use of an object. Second, the solution may be further impeded because the first solution attempt often causes the solver to fall into a rut. Gestalt psychologists have called the inability to escape a mental set, such as the one associated with an initial solution attempt, *fixation* (Woodworth & Schlosberg, 1954). Thus, it is not only the inappropriateness of the initial mental set but, more importantly, the fixation promoted by such a set that delays or blocks a successful solution to these problems.

More recently, fixation has been studied in relation to conformity with examples. For instance, Jansson and Smith (1991) found that design students produced inappropriate solutions on a design task more often after being shown an inappropriate example. Smith, Ward, and Schumacher (1993) found that showing subjects an example constrained the generation of new category members. Similarly, students' knowledge of real animals constrained their generation of imaginary animals (Ward, 1994). In these cases, the example seemed to establish a mental set, which, in turn, promoted fixation on features of the example.

In another recent demonstration, S. M. Smith and Blankenship (1991) examined fixation using Mednick's remote associate task (RAT). The RAT grew out of Mednick's (1962) belief that the novel combination of remote associations underlies the creative process. The standard RAT

involves the presentation of three words, such as *blue*, *knife*, and *cottage*, and prompts the subject to generate a fourth word that forms a familiar phrase with each of three. In this case, the fourth word could be *cheese*, forming the phrases *blue cheese*, *cheese knife*, and *cottage cheese*. S. M. Smith and Blankenship induced fixation on the RAT by priming meanings of words that were irrelevant to solution. When an irrelevant associate was presented in italics next to each word in a problem, the subjects were significantly less likely to solve the problems. However, when the subjects were given an incubation period between the initial problem-solving attempt (with the irrelevant prime) and a second attempt (without primes), the effects of the primes dissipated, and the subjects' ability to solve the problems improved. One way of viewing these findings is that priming the irrelevant meanings instilled a mental set, which produced an inappropriate initial problem-solving attempt as well as fixation on the incorrect solutions, preventing a broad search of the solution space.

The present study investigates whether domain knowledge may be seen to influence solution attempts in a similar way. Since domain knowledge is easily and possibly automatically activated when experts encounter domain-relevant material, domain knowledge may have an influence on solution attempts that is similar to that of the externally imposed primes in the studies above. In other words, domain knowledge may act as a mental set and promote fixation in problem solving. Although it may not be surprising to find that expert knowledge can act as a mental set, as experts may rely on their domain knowledge to inform an initial solution attempt, the more interesting question is whether experts will experience more fixation than novices will, when the initial solution that is suggested by domain knowledge is inappropriate. Of further interest is how such fixation might be overcome and whether the source of fixation plays a role in whether or not an incubation period will improve problem solving.

In three experiments, the possible existence and nature of a mental set and of fixation that is due to the possession of a large amount of domain knowledge was investigated. The first experiment demonstrated that domain knowledge can indeed act as a mental set and promote fixation in creative problem solving. In the second experiment, it was investigated whether the fixation that was due to domain knowledge could be eliminated with a warning. In the third experiment, the effect of incubation on fixation was examined as a function of whether the fixation was internally generated (knowledge based), rather than externally primed.

EXPERIMENT 1

To test the hypothesis that domain knowledge can act as a mental set and promote fixation in creative problem solving, subjects with varying amounts of baseball knowledge performed an RAT (adapted from Mednick, 1962). For this experiment, 10 items were created, such that the

first word could have been part of a baseball-related term. For example, 1 baseball-related problem consisted of the words *plate*, *broken*, and *rest*, with the intended solution of *home*, which would form the phrases *home plate*, *broken home*, and *rest home*. Another 10 problems were based on Mednick's original stimuli. As none of the Mednick-based items contained a first word that was related to a baseball term, these trials are termed *neutral*. The *blue*, *knife*, *cottage* example that was mentioned above is an example of a neutral trial.

Two versions were created for each baseball-related trial. In one version—for example, the *plate*, *broken*, *rest* example given above—the answer that was suggested by the baseball-related meaning of the first word formed a phrase with both the second and the third words and thus was the solution. These versions of the trials are termed *baseball consistent*. For a second version of each trial, the third word was replaced by a word that would not form a good phrase with the answer suggested by the baseball-related meaning of the first word. For example, the second version of the above problem had the words *plate*, *broken*, and *shot*. The intended answer was *glass*, forming the phrases *plate glass*, *broken glass*, and *shot glass*. In these trials, the answer suggested by the baseball-related meaning of the first word formed a phrase with the second word but not with the third. Because the subject did not know that the solution suggested by the baseball term would not work until the third word appeared, these versions of the baseball-related trials were termed *baseball misleading*.

If expertise does impose a set and promote fixation in problem solving, subjects with higher amounts of baseball knowledge should be less able to generate correct solutions than subjects with less domain knowledge when answers primed by baseball terms are incorrect (that is, on baseball-misleading trials). It is presumed that there will be no differences in performance on neutral trials.

Method

Subjects. Twenty-four University of Pittsburgh undergraduates enrolled in introductory psychology courses participated in this experiment to fulfill a course requirement. Subjects were run in groups of 1 to 12.

Materials. Mednick's (1962) RAT was adapted for use in this experiment. The adapted RAT was composed of 20 items. Ten of the items were based on Mednick's original stimuli. Another 10 items were created with a first word that could be part of a baseball-related phrase. The second and third words in all items were neutral—that is, not baseball related. Two versions were made of each baseball-related item. In one version, the trials were baseball-consistent, and the solution was the word that formed a baseball-related phrase with the first term. In a second version, the trials were baseball-misleading, and the solution was not the word that formed a baseball-related phrase with the first term. To create this second version, the third word of each baseball trial was replaced by another word. The baseball-misleading trials did have a solution, just not the one suggested by the baseball-related meaning of the first term. The stimuli for the adapted RAT are presented in the Appendix.

Procedure. The subjects were told that they would perform several problem-solving tasks. The experiment was run on IBM-PC compatible computers using Micro Experimental Lab software

Table 1
Mean Scores (and Standard Deviations) on
Domain Knowledge Questionnaire and Gender Breakdown for Each Experiment

Experiment	High Knowledge				Low Knowledge			
	<i>M</i>	<i>SD</i>	No. Males	No. Females	<i>M</i>	<i>SD</i>	No. Males	No. Females
1	23.4	6.7	9	3	8.4	4.7	4	8
2	26.4	7.7			7.2	2.6		
No warning			5				1	4
Warning			4	1			2	3
3	24.7	5.0			8.9	3.5		
No fixation, no incubation			4	1			1	4
No fixation, incubation			3	2			3	2
Fixation, no incubation			3	2			1	4
Fixation, incubation			3	2			1	4

(Schneider, 1988). Before beginning the RAT, the subjects were given an example, five practice items, and some acceptable solutions for the practice items. All the subjects then performed the adapted RAT. In this version of the task, the three words were presented one at a time, cumulatively. This presentation format differs from both Mednick's (1962) and S. M. Smith and Blankenship's (1991), and was instituted to maximize the possibility of finding fixation effects. The first word appeared alone in the center of the screen for 5 sec, after which it was joined by the second word, which appeared under it. After another 7.5 sec, the first two words were joined by the third word, which appeared under them, and the subject had 30 sec in which to produce the fourth word. The subject could answer at any time during the 30-sec problem-solving period by pressing the space bar, typing in a solution word, and pressing the return key. However, if the 30 sec passed and the subject had not yet responded, the three words disappeared, and the subject was told to enter a solution word and press the return key. The subject had 20 sec to respond, after which the next trial was presented.

Each subject received 10 neutral trials and 10 baseball-related trials. Five of the baseball-related trials were consistent with the solution suggested by the baseball term, while 5 were misleading. Presentation order and baseball trial type were randomized, such that each subject received only one version of each item. Responses and response times were recorded by computer. The response time that was analyzed was the duration from the onset of the third word until the subject finished entering a word and pressed the return key or until the trial ended (including both the 30-sec problem-solving and 20-sec response periods).

The subjects then completed a 24-item mental rotation task (Shepard & Metzler, 1971), a 10-item anagram task, a 9-item Gestalt completion task (Street, 1931), and a multiple-choice baseball term recognition test. In the recognition task, the subjects were asked to pick the word that completed the baseball term—for example, *word*: (1) time, (2) run, (3) house, (4) series. All four additional tasks were presented by computer. The subjects were given up to 15 sec to complete each mental rotation task, 20 sec to complete each anagram, and 30 sec to complete each Gestalt completion task. The baseball recognition task had no time constraint. Responses and response times were recorded by computer. The final task was a 45-item paper-and-pencil short-answer test of baseball knowledge, the Baseball Knowledge Questionnaire from Spilich, Vesonder, Chiesi, and Voss (1979). No feedback was given on any task.

Scoring. Because a number of RAT problems received responses that were not the intended solution, an acceptability rating procedure was devised. First, two judges categorized all solutions as either *definitely not correct* or *possibly correct*. Items that both judges called *definitely not correct* were coded as incorrect. For instance, the answer *black* to the item *sleep*, *board*, and *cat* was coded as incorrect because, although *black board* and *black cat* are good

phrases, both judges agreed that *black sleep* or *sleep black* was not. Answers that one or both of the judges categorized as possibly correct responses ($n = 71$), as well as the intended answers ($n = 30$), were then given to a new group of subjects to be rated for acceptability. Ten raters (5 high-knowledge and 5 low-knowledge) were shown the same example item and practice items with suggested solutions as the experimental subjects. They were then asked to rate on a 10-point scale how *good* of a phrase each possible solution made with each of the words in each problem, with 1 meaning *not at all good*, and 10 meaning *very good*. All the solutions that formed phrases receiving average ratings similar to the intended answers, either in overall rating or on a phrase-by-phrase basis, were accepted as correct solutions. To be specific, an overall rating for each answer was computed by summing the goodness ratings for that answer with all three words in a problem. The lowest average overall rating for an intended answer was 23.8 ($SD = 3.3$). All unintended answers with average overall ratings greater than 23.8 were accepted as correct ($n = 8$). In addition, the lowest per phrase average goodness ratings for intended answers were 7.2, 5.1, and 7.0, for the first, second, and third words of a problem, respectively. All unintended answers that had an average overall rating greater than 19.3 but no single rating less than 5.1 were also accepted as correct ($n = 9$).

The mental rotation, Gestalt completion, and anagram tasks served as filler tasks in later experiments. The recognition task was employed simply to ensure that all the subjects, even low-knowledge subjects, were familiar with the baseball terms. The Baseball Knowledge Questionnaire was used to determine the relative knowledge level of the subjects.

Results

The subjects were divided into two baseball knowledge groups via median split on the scores of the Baseball Knowledge Questionnaire. Performance on the RAT was analyzed in terms of proportion of correct solutions and response time for each knowledge group and trial type. For all analyses, analyses of variance (ANOVAs) were computed, with repeated measures on trial type. Both significant and nonsignificant main effects are reported. However, interactions are only reported when they are significant, unless they pertain to an experimental hypothesis. Especially of interest was the presence of a *misleading effect*, or decrement in performance on misleading trials, for the high-knowledge subjects.

Baseball knowledge. Performance on the Baseball Knowledge Questionnaire ranged from 2 to 32 correct answers out of a possible 45. Those subjects scoring less than

Table 2
Mean Proportions (and Standard Deviations) of Correct Solutions,
Intrusions, and Response Times (in Seconds) in Experiment 1

Condition	Solutions		Intrusions		RT	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Low knowledge						
Neutral	.50	.29			20.6	11.5
Consistent	.40	.28			20.9	12.4
Misleading	.40	.16	.11	.15	23.0	12.2
High knowledge						
Neutral	.50	.18			24.7	13.4
Consistent	.38	.18			26.8	14.3
Misleading	.15	.17	.31	.37	31.4	15.2

15 were categorized as low knowledge, and subjects scoring 15 or more were categorized as high knowledge. The mean scores and gender breakdown of each knowledge group are included in Table 1. Even the low-knowledge subjects were familiar with the baseball terms used in this study, as all the subjects were able to indicate the correct baseball terms on the baseball term recognition test.

The high-knowledge groups in these experiments have noticeably lower means on the Baseball Knowledge Questionnaire than do the high-knowledge groups in the Voss studies, since they selected subjects after pretesting large groups on the questionnaire. For example, the high-knowledge groups in Voss et al. (1980) had means of 41.3 and 40.0. Although the lower mean allowed for a mix of male and female subjects among the high-knowledge groups, it creates a concern that the observed effects may be more appropriately attributed to intermediate levels of domain knowledge. A separate analysis performed on just the highest knowledge subjects in the present study (who were comparable to the high-knowledge subjects in previous studies) confirmed that significant misleading effects could be found in the most knowledgeable subjects as well. Because this analysis pooled across the three experiments, it is included at the end of the third experiment.

Correct solutions. A 2×3 (knowledge level \times trial type) ANOVA was computed, using proportion of correct solutions as a dependent measure. Table 2 presents the mean proportion of correct solutions for each condition. No significant main effect was found for knowledge level, but there was a significant effect of trial type [$F(2,44) = 8.56$, $MS_e = 0.03$, $p < .001$]. Tukey's HSD indicated that there were more correct solutions to neutral trials than to misleading trials, with consistent trials not differing significantly from either of the other two.

There was a significant interaction between knowledge level and trial type [$F(2,44) = 4.06$, $MS_e = 0.03$, $p < .02$]. Pairwise comparisons revealed that, for high-knowledge subjects, misleading items were solved less often than either neutral [$t(11) = 4.69$, $p < .001$] or consistent items [$t(11) = 3.09$, $p < .01$]. There was no difference in the number of correct solutions between neutral and consistent items for the high-knowledge subjects ($t = 1.66$), nor

between any of the items for the low-knowledge subjects ($t_s < 1.3$).

A contrast comparing correct solutions for neutral and consistent trials together against misleading trials revealed a significant misleading effect for the high-knowledge subjects [$t(11) = 4.53$, $p < .0001$] but not for the low-knowledge subjects ($t < 1$). Further, the interaction of this misleading effect by knowledge group was significant [$F(1,44) = 4.03$, $MS_e = 0.03$, $p < .02$ —that is, high-knowledge subjects were more likely to show a decrement on misleading problems versus neutral and consistent problems than were low-knowledge subjects. This suggests that the knowledge \times trial type interaction was due to a greater misleading effect among the high-knowledge subjects.

Incorrect answers on misleading trials were further analyzed on the basis of whether they were related to baseball. Incorrect answers related to baseball were termed *intrusions*. High-knowledge subjects made baseball-related intrusions on .31 ($SD = .37$) of misleading trials, whereas low-knowledge subjects only made baseball-related intrusions on .11 ($SD = .15$) of the misleading trials. This difference approached significance [$t(22) = 1.72$, $p < .10$], suggesting that the more knowledgeable subjects tended to be fixated in the solution suggested by their baseball knowledge.

Response time. A 2×3 (knowledge level \times trial type) ANOVA that used response time as a dependent measure was computed. As can be seen in Table 2, there was a significant main effect of trial type [$F(2,44) = 4.68$, $MS_e = 28.16$, $p < .01$]. Tukey's HSD indicated that misleading trials had significantly longer response times than either neutral or consistent trials, which were not significantly different from each other. The main effect for knowledge level was not significant ($F = 1.45$), nor was the interaction between knowledge level and trial type ($F = 1.00$), although response times did tend to be longer for the high-knowledge subjects on the misleading trials.

Discussion

The misleading effect was greater for high-knowledge subjects than for low-knowledge subjects. For misleading items, high-knowledge subjects were much less likely

than were low-knowledge subjects to arrive at a correct solution. High-knowledge subjects were also more likely to give incorrect baseball-related responses to misleading items. These results suggest that high-knowledge subjects were fixated in their problem solving.

One explanation for this effect is that the domain knowledge of experts is being activated by the content of the problems. Domain knowledge not only acts as a mental set suggesting an initial solution, but also fixates the high-knowledge subject on the domain-related initial solution attempt. This is consistent with previous findings (see, e.g., Ward, 1994) that knowledge of real-world examples can constrain the generation of novel animals or objects. Further, the present study extends such findings by investigating fixation effects among problem solvers with different levels of knowledge.

However, an alternative explanation is that high-knowledge subjects may simply be more sensitive to the fact that a number of the trials seem to do with baseball. Because using one's domain knowledge is usually of benefit, experts may be engaging their knowledge intentionally, without realizing that it is harming their performance on the misleading trials (note that there was no feedback given). Thus, it seems important to differentiate fixation that might be occurring as the result of an intentional strategy ("Hey, a lot of these are about baseball, I should look for a baseball-related term") from fixation that is the result of the more-or-less automatic activation of the knowledge structure. This issue is the focus of Experiment 2.

A concern that may be raised because of the use of a median-split procedure is simply that the high-knowledge and low-knowledge populations may vary in more than just their knowledge of baseball and that ability differences between the two groups may be underlying the poorer problem solving in the high-knowledge group. However, this concern does not seem warranted here. To begin with, the critical variable is a within-subjects difference score—that is, high-knowledge subjects perform as well as low-knowledge subjects on neutral trials but less well on the misleading trials. Second, previous studies using median-split procedures on the same Baseball Knowledge Questionnaire have looked directly at measures of ability and found no differences between high- and low-knowledge groups (see, e.g., Fincher-Kiefer, Post, Greene, & Voss, 1988; Spilich et al., 1979).

It is interesting that no facilitating effects were found for the high-knowledge subjects on the baseball-consistent trials. One possible explanation is that the baseball-related solutions were easily accessible for both the high- and low-knowledge subjects; hence, low-knowledge subjects were able to solve the baseball-consistent problems just as well as the high-knowledge subjects. On the other hand, low-knowledge subjects were able to ignore the baseball-related meaning when it was not appropriate, whereas high-knowledge subjects were not, which resulted in fixa-

tion among the high-knowledge subjects on the misleading problems.

Another possibility is that, if the subjects had been given just baseball-consistent problems, high-knowledge subjects might have shown an advantage over the low-knowledge subjects. It may be that the presence of the baseball-misleading items in the present experiment caused interference even on the baseball-consistent problems.

Because the main interest of this study is the fixation that may be caused by domain knowledge and because the presence of the baseball-consistent trials may be cueing the use of baseball knowledge, for Experiments 2 and 3 only neutral and misleading trials were used.

EXPERIMENT 2

Experiment 1 demonstrated mental set and fixation in problem solving among high-knowledge subjects. Experiment 2 tested whether this fixation that was due to domain knowledge could be eliminated by warning the subjects not to use their baseball knowledge. The subjects either received no information about the task, as in Experiment 1, or were told explicitly that some of the items might be related to baseball terms but that thinking of such terms would not help their performance. Further, to eliminate any helpfulness of baseball knowledge, the subjects were given only misleading baseball trials.

If the mental set and fixation of high-knowledge subjects is due to the assumption that it is beneficial to use one's knowledge whenever it seems relevant, high-knowledge subjects should be more successful at solving misleading trials when they are explicitly told that their knowledge will not help them. However, if, as predicted, the mental set and fixation of high-knowledge subjects is related to the activation of knowledge itself, warning high-knowledge subjects not to use their knowledge should not produce any improvement in performance on misleading trials, as compared with a no-warning condition.

Method

Subjects. Ten University of Pittsburgh undergraduates enrolled in introductory psychology courses participated in this experiment to fulfill a course requirement. In addition, 10 volunteers from the university community who responded to newspaper, poster, and usenet advertisements participated in this experiment and received \$5. The subjects were run in groups of 1 to 6. Both the warning and nonwarning conditions consisted of an equal number of paid and unpaid subjects.

Because domain knowledge was not assessed until the end of the experiment, 4 additional subjects were run to complete the design. In all cases, only the first 5 subjects run in each cell were used for analysis, yielding an equal number of high-knowledge and low-knowledge subjects in each condition.

Procedure. The procedure was the same as that in Experiment 1, except that, after the possible solutions to the practice trials were presented, half the subjects received the warning, "Some of these items may bring baseball-related terms to mind. However, you should try to ignore them because they will *not* lead you to the correct so-

Table 3
Mean Proportions (and Standard Deviations) of Correct Solutions,
Intrusions, and Response Times (in Seconds) in Experiment 2

Condition	No Warning						Warning					
	Correct		Intrusions		RT		Correct		Intrusions		RT	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Low knowledge												
Neutral	.56	.15			23.1	2.2	.38	.26			26.6	10.4
Misleading	.46	.09	.15	.10	26.5	3.2	.26	.18	.14	.10	26.8	8.1
<i>Effect</i>	.10				3.4		.12				0.2	
High knowledge												
Neutral	.58	.08			18.5	3.1	.52	.16			22.0	9.6
Misleading	.20	.16	.24	.21	27.9	3.6	.30	.14	.18	.06	29.9	6.2
<i>Effect</i>	.38				9.4		.22				7.9	

Note—Misleading Effect for Correct Solutions = (Proportion correct on neutral items) – (Proportion correct on misleading items). Misleading Effect for Response Time = (Response time on misleading items) – (Response time on neutral items).

lution.” The other half of the subjects received no warning, as in Experiment 1. All the subjects received 10 misleading baseball trials and 10 neutral trials. Item presentation order was randomized for each subject.

Results

Performance on the RAT was analyzed in terms of proportion of correct solutions and response time for each condition. Especially of interest were the replication of the misleading effect found for the high-knowledge subjects in Experiment 1 and the effect of the warning manipulation.

Baseball knowledge. Performance on the Baseball Knowledge Questionnaire ranged from 3 to 39 correct answers, with those subjects scoring less than 15 categorized as low-knowledge and those subjects scoring 15 or more categorized as high-knowledge. The mean scores and gender breakdown for each condition and knowledge group are included in Table 1. All the subjects were able to indicate the correct baseball terms on the baseball term recognition test.

Correct solutions. Using the same procedure as that in Experiment 1, a $2 \times 2 \times 2$ (knowledge level \times warning \times trial type) ANOVA was computed on the number of correct solutions. Table 3 presents the mean proportion correct for each condition. No significant main effects were found for knowledge level or warning. There was a significant main effect for trial type [$F(1,16) = 19.66$, $MS_e = 0.02$, $p < .001$], as more correct solutions were offered for neutral than for misleading trials. The interaction between knowledge level and trial type was marginally significant [$F(1,16) = 4.22$, $MS_e = 0.02$, $p < .06$].

On the basis of the results of Experiment 1, a planned multiple comparison was computed on the no-warning condition alone. It indicated that, without the warning, high-knowledge subjects were less likely than low-knowledge subjects to make a correct response on misleading trials versus neutral trials [$F(1,8) = 8.34$, $MS_e = 0.01$, $p < .02$], replicating Experiment 1. Planned pairwise comparisons within each knowledge group revealed a significant misleading effect for the high-knowledge sub-

jects [$t(4) = 7.76$, $p < .01$] but none for the low-knowledge subjects [$t(4) = 1.20$, $p > .29$].

Interestingly, although the main effect for warning was not significant [$F(1,16) = 2.32$, $MS_e = 0.02$, $p < .15$], the manipulation appeared to have the opposite of its intended effect, with a greater proportion of correct solutions overall in the no-warning condition ($M = .46$) than in the warning condition ($M = .37$). Although the warning was supposed to help the experts, it seemed instead to hurt the performance of low-knowledge subjects, as the knowledge level \times warning condition interaction was marginally significant [$F(1,16) = 3.54$, $MS_e = 0.02$, $p < .08$]. A post hoc comparison for the low-knowledge subjects, using Scheffe's test, indicated that the low-knowledge subjects made fewer correct responses on both neutral and misleading items in the warning condition than in the no-warning condition [$F(1,16) = 8.44$, $MS_e = 0.02$, $p < .01$]. The same test on the performance of high-knowledge subjects yielded no difference between warning and no-warning conditions ($F < 0.1$).

Although the three-way (knowledge level \times warning \times trial type) interaction was not significant, a planned comparison computed for just the high-knowledge subjects revealed that a warning did not significantly improve performance on misleading versus neutral items for the high-knowledge subjects [$F(1,16) = 1.89$, $MS_e = 0.02$, $p > .10$], indicating that a warning did not lessen the misleading effect for the high-knowledge subjects. Taken together, these results suggest that, rather than selectively helping the high-knowledge subjects to overcome their fixation on misleading items, the warning manipulation may have caused fixation on all items to some extent for the low-knowledge subjects.

Finally, as in Experiment 1, high-knowledge subjects tended to offer a greater number of incorrect baseball-related solutions on the misleading trials (.21, $SD = .15$), or intrusions, than did the low-knowledge subjects (.14, $SD = .09$), although this result was not significant [$t(18) = 1.24$, $p < .23$].

Response time. A $2 \times 2 \times 2$ (knowledge level \times warning \times trial type) ANOVA, using response time as a

dependent measure, was computed. The means for each condition are presented in Table 3. There was no significant effect of knowledge level on response time, nor was there a significant main effect that was due to the warning. A significant main effect was observed for trial type [$F(1,16) = 19.93$, $MS_e = 13.73$, $p < .001$], with misleading trials taking longer than neutral trials. Further, the overall knowledge level \times trial type interaction was significant, indicating that, regardless of warning condition, the high-knowledge subjects showed a significantly greater slowdown on the misleading trials, as compared with neutral trials, than did the low-knowledge subjects [$F(1,16) = 8.74$, $MS_e = 13.73$, $p < .01$].

Discussion

Experiment 2 replicated the main result of Experiment 1—namely, that, on misleading trials, unwarned high-knowledge subjects were less likely than unwarned low-knowledge subjects to arrive at a correct solution, indicating that the high-knowledge subjects were fixated in their problem-solving attempts. Further, although there did seem to be a slight improvement on misleading items, warning the high-knowledge subjects not to use their domain knowledge did not significantly improve their performance.

The warning manipulation seems to have had an unintended effect on the low-knowledge subjects. The decrease in the proportion of correct solutions for the low-knowledge subjects on all trials in the warning condition suggests that the low-knowledge subjects were fixated in some way by the instruction not to think of baseball-related solutions. The suggestion not to use one's baseball knowledge may have compelled the low-knowledge subjects to look for baseball-related solutions, similar to the effects of suggestions not to think of white bears or pink elephants that Wegner (1989) has observed. Unfortunately, the subjects were not asked whether they actually believed the warning, and the subjects could also have been driven to look for baseball-related solutions if they felt they were being deceived by the experimenter.

Although it is unclear whether the warning failed to significantly improve performance because subjects were unable to stop thinking about baseball or because subjects did not believe the warning, these results are consistent with a number of experiments that have found that people have difficulty avoiding the use of stored or presented examples in creativity tasks, even when they are warned not to use such information (Jansson & S. M. Smith, 1991; S. M. Smith & Tindell, 1997; S. M. Smith et al., 1993; Ward, 1994). Further, this experiment did provide a clear replication of Experiment 1. In the nonwarning condition, along with an observed misleading effect on the proportion of correct solutions, high-knowledge subjects had longer response times on misleading versus neutral problems than did low-knowledge subjects. Also, as in Experiment 1, the high-knowledge subjects were more likely to make intrusions (incorrect baseball-related solutions) on the misleading items than were the low-knowledge sub-

jects. These results offer converging evidence that the high-knowledge subjects were fixated in their solution attempts by the activation of their domain knowledge.

EXPERIMENT 3

An important distinction should be made between the mental set and fixation that are due to domain knowledge, observed in the present study, and the kinds of mental sets that have been considered in the literature, for the most part, that are experimentally produced. Traditional demonstrations of mental set and fixation have used sets that are externally imposed on the subject through the presentation of an object in a specific context or through practice at a task prior to the experimental task. On the other hand, the proposed fixation from expertise is the result of an internally generated set, and this may lead to some interesting differences in how the fixation that is due to extensive domain knowledge may or may not be overcome.

Most prominently, it would seem that, if the fixation observed in the high-knowledge individuals is the result of the easy activation of a well-formed knowledge base, simply introducing an *incubation* period may not do anything to help free a subject from the fixation that is due to an internally generated set. However, an incubation period may help break fixation from an externally imposed set, as the delay would allow for the dissipation of the inappropriate activation. This would, in turn, suggest that *distributed* effort toward problem solution (i.e., taking a break during the solution phase) may be the most effective tactic for less knowledgeable subjects. More knowledgeable subjects, on the other hand, should not benefit from a distributed effort, since their success depends on the dismissing of internally activated irrelevant stimuli. Even after a delay, high-knowledge subjects should experience the same fixation when their domain knowledge is activated by the content of the reintroduced problem.

In a third experiment, the extent to which the mental set and fixation that are a function of domain knowledge are different from an externally imposed set was investigated. As noted previously, S. M. Smith and Blankenship (1991) have demonstrated that fixation can be induced by the presentation of irrelevant word associations before or during the RAT. They have further shown that the fixation that is due to such priming dissipates after an incubation period, presumably as the irrelevant activation fades. The fixation that is due to internally generated sets, however, may not be as easy to overcome, since the activation of *irrelevant* word meanings comes from the expert's domain knowledge. To test this notion, the first experiment was repeated, but this time fixation was induced. One half of the high- and low-knowledge subjects were primed for the misleading meanings of terms included in the RAT (i.e., baseball terms for the baseball trials and irrelevant associations for the neutral trials). Further, all the subjects performed the RAT twice—half immediately after the first attempt and half with a 10-min period of incubation intervening, during which the

Table 4
Mean Proportions (and Standard Deviations) of Correct Solutions,
Intrusions, and Response Times in Experiment 3

Condition	No Fixation Induced						Fixation Induced					
	Correct		Intrusions		RT		Correct		Intrusions		RT	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Low knowledge												
Neutral	.53	.24			21.7	6.4	.58	.16			19.3	5.7
Misleading	.46	.19	.14	.10	25.7	10.1	.32	.15	.30	.17	24.8	8.2
<i>Effect</i>	.07				4.0		.26				5.5	
High knowledge												
Neutral	.64	.16			17.8	4.8	.58	.19			18.5	6.4
Misleading	.27	.15	.20	.12	27.8	7.1	.36	.21	.33	.30	23.7	6.8
<i>Effect</i>	.37				10.0		.22				5.2	

Note—Misleading Effect for Correct Solutions = (Proportion correct on neutral items) – (Proportion correct on misleading items). Misleading Effect for Response Time = (Response time on misleading items) – (Response time on neutral items).

subjects performed demanding tasks. It was predicted that low-knowledge subjects would show an effect due to the fixation manipulation, as well as more of an incubation effect, as compared with high-knowledge subjects. High-knowledge subjects should show neither an effect due to the fixation manipulation nor any improvement due to incubation.

Method

Subjects. Forty volunteers from the university community who responded to newspaper, poster, and usenet advertisements participated in this experiment and received \$5. The subjects were run in groups of 1 to 4. Because baseball knowledge was not assessed until the end of the experiment, an additional 3 subjects were run to complete the design. In all cases, only the first 5 subjects run in each cell were used for analysis, yielding equal numbers of high- and low-knowledge subjects in each condition.

Procedure. The procedure was the same as that in Experiment 1, except that half of the subjects performed a *short-term memory task* prior to the experiment. This was the *fixation-induced* condition. The other half of the subjects did not perform the short-term memory task and began the experiment with the RAT, as in Experiment 1. This was the *nonfixation* condition. In addition, all the subjects received the RAT twice. Incubation was manipulated by having half of the subjects perform the second RAT immediately after the first (*no-incubation* condition), whereas the other half received the second RAT after the mental rotation, anagram, and Gestalt completion tasks (*incubation* condition). The subjects in the no-incubation condition performed these tasks after the second RAT. All the subjects then completed the baseball term recognition test and the Baseball Knowledge Questionnaire.

The task that was used to induce fixation was based on a procedure developed by S. M. Smith and Schumacher (1992). The subjects were told that they were performing a *short-term memory task*. Three words were presented on the screen at the same time. Subjects had 5 sec to find two 2-word phrases that could be made out of the words. They then had to remember those phrases for 10 sec, after which they were asked to type in their responses. The words used for this task were two words from each RAT item and either the baseball-related solution (for the baseball trials) or a word that formed phrases with the words but was not the solution (for the neutral trials). The stimuli that were used for each item are included in the Appendix in parentheses. The actual presentation order of the words was such that no two words in succession formed a phrase. The order of the items was randomized for each subject.

At no time in this experiment were the subjects told that the study could be related to baseball.

For both administrations of the RAT, all the subjects received 10 neutral trials and 10 misleading baseball trials.

Results

Performance on the first RAT was analyzed in terms of proportion of correct solutions and response time for each knowledge level and fixation group. In addition, performance on the second administration of the RAT was analyzed in terms of improvement over the first administration as a function of knowledge group, fixation, and incubation group. Of specific interest were replicating the misleading effects found for high-knowledge subjects in earlier experiments, testing whether a misleading effect could be induced in low-knowledge subjects by the fixation manipulation, and, if so, testing whether it could then be attenuated by incubation.

Baseball knowledge. Performance on the Baseball Knowledge Questionnaire ranged from 3 to 33 correct answers, with those subjects scoring less than 15 categorized as low knowledge and those subjects scoring 15 or more categorized as high knowledge. The mean scores and gender breakdown for each condition are shown in Table 1. All the subjects indicated the correct baseball terms on the recognition test.

Correct solutions on the first remote associates task. Using the same *acceptability* rating procedure for correct solutions as that in Experiments 1 and 2, a $2 \times 2 \times 2$ (knowledge level \times fixation \times trial type) ANOVA was computed on the number of correct solutions. Table 4 presents the mean proportion correct for each condition. No significant main effects were found for either knowledge level or fixation. There was a significant main effect for trial type [$F(1,36) = 42.53$, $MS_e = 0.03$, $p < .0001$], as a greater number of correct solutions were offered for neutral trials than for misleading trials. Further, the three-way interaction between fixation condition, knowledge level, and trial type was significant [$F(1,36) = 4.93$, $MS_e = 0.03$, $p < .03$]. Planned multiple comparisons were performed for each fixation condition and knowledge level. As was found in Experiments 1 and 2, a planned multiple comparison in the nonfixation condition indicated that high-knowledge subjects were less likely to make a correct

Table 5
Mean Proportions (and Standard Deviations) of
Unsolved Items Solved at Retest in Experiment 3

Condition	Low Knowledge				Effect	High Knowledge				Effect
	No incubation		Incubation			No incubation		Incubation		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
No fixation										
Neutral	.32	.27	.40	.40	.08	.56	.29	.21	.22	-.35
Misleading	.07	.15	.21	.29	.14	.12	.16	.11	.07	-.01
Fixation										
Neutral	.20	.22	.35	.27	.15	.50	.29	.29	.10	-.21
Misleading	.11	.17	.27	.29	.16	.30	.22	.09	.12	-.21

Note—Incubation Effect = (Incubation improvement) – (No incubation improvement).

response on misleading trials, as compared with neutral trials, than were low-knowledge subjects [$F(1,18) = 6.36$, $MS_e = 0.03$, $p < .02$]. Pairwise comparisons within knowledge group revealed a significant misleading effect for the high-knowledge subjects [$t(9) = 4.69$, $p < .001$] but none for the low-knowledge subjects [$t(9) = 1.15$, $p > .27$]. The same overall multiple comparison did not approach significance in the fixation-induced condition ($F < 1$).

Planned pairwise comparisons were performed to test the effects of the fixation manipulation for each knowledge group. The fixation manipulation did have its intended effect among the low-knowledge subjects, as they made fewer correct responses on misleading versus neutral trials in the fixation condition [$t(9) = 9.75$, $p < .0001$]. High-knowledge subjects made fewer correct responses on misleading versus neutral items in the fixation condition as well [$t(9) = 2.50$, $p < .03$]. Thus, both high-knowledge and low-knowledge subjects showed a misleading effect in the fixation condition.

Further, the fixation manipulation significantly increased the proportion of intrusions on misleading trials among both the low- and high-knowledge subjects [$t(38) = 2.57$, $p < .01$]. The mean proportions of intrusions for each condition and their standard deviations are presented in Table 4. As in the first two experiments, in the nonfixation conditions, high-knowledge subjects tended to make more intrusions overall (.20, $SD = .12$) than did low-knowledge subjects (.14, $SD = .10$) but not significantly so [$t(18) = 1.21$, $p < .24$]. Taken together, these results suggest that the fixation-inducing procedure used in Experiment 3 was successful at inducing a set in the low-knowledge subjects that was similar to that of the high-knowledge subjects.

Response time on the first remote associates task. A $2 \times 2 \times 2$ (knowledge level \times fixation \times trial type) ANOVA, using response time as a dependent measure, was computed. The means for each condition are presented in Table 4. There was no main effect for either knowledge level or fixation. A significant main effect was observed for trial type [$F(1,36) = 30.50$, $MS_e = 24.98$, $p < .0001$], with misleading trials taking longer than neutral trials. The three-way interaction between fixation, knowledge level, and trial type approached significance [$F(1,36) = 2.72$, $MS_e = 24.98$, $p < .11$]. On the basis of the results

of Experiment 2, a multiple comparison in the nonfixation condition indicated that high-knowledge subjects tended to have longer response times on misleading versus neutral trials than did low-knowledge subjects [$F(1,18) = 3.45$, $MS_e = 26.71$, $p < .08$], which is similar to the result from Experiment 2.

Improvement on second remote associates task. The measure of improvement from the first RAT to the second was taken from S. M. Smith and Blankenship (1991): improvement = (number newly solved at retest) / (total number of problems – number solved at first test). This measure represents the proportion of problems not solved during the initial test that were correctly solved at retest. A $2 \times 2 \times 2 \times 2$ ANOVA (knowledge level \times fixation \times incubation \times trial type) was computed, using this measure. There was a significant effect of trial type [$F(1,32) = 15.22$, $MS_e = 0.05$, $p < .001$]—that is, at retest, more of the unsolved neutral problems were solved than of the unsolved misleading problems. Further, a two-way interaction between knowledge level and incubation was significant [$F(1,32) = 8.59$, $MS_e = 0.05$, $p < .01$]. As is illustrated in the third and last column of Table 5, low-knowledge subjects solved more problems at retest when there was a delay (incubation period) between their initial test and retest than when the retest immediately followed the initial test. In other words, low-knowledge subjects demonstrated a positive incubation effect. On the other hand, high-knowledge subjects solved more unsolved problems at retest when the retest immediately followed the initial test than when there was a delay, resulting in negative incubation effects.

Discussion

The nonfixation condition replicated the results of Experiments 1 and 2, as high-knowledge subjects were significantly less likely than low-knowledge subjects to produce a correct response to misleading trials in the first RAT, demonstrating a mental set and fixation that were due to domain knowledge. High-knowledge subjects also tended to have longer response times on misleading trials than did low-knowledge subjects, which is similar to the finding from Experiment 2.

Low-knowledge subjects showed a significant misleading effect that was due to the fixation manipulation,

Table 6
Misleading Effects (and Standard Deviations)
Across All Three Experiments

Condition	Dependent					
	Correct Solutions		Intrusions		RT	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Low knowledge						
Neutral	.52	.24			21.5	8.5
Misleading	.43	.16	.13	.12	24.6	10.1
<i>Effect</i>	.09				3.8	
High knowledge						
Neutral	.57	.17			21.0	9.9
Misleading	.20	.17	.26	.27	29.4	10.1
<i>Effect</i>	.37				8.4	

Note—Misleading Effect for Correct Solutions = (Proportion for neutral items) – (Proportion for misleading items). Misleading Effect for Response Time = (Response time on misleading items) – (Response time on neutral items).

as they made fewer correct solutions on misleading items than on neutral items when they had been primed on irrelevant solutions. High-knowledge subjects, on the other hand, performed worse on misleading items, regardless of fixation condition. This finding is consistent with the idea that the expert's mental set may be the result of the automatic activation of their domain knowledge. For the high-knowledge subjects, the priming of the irrelevant solutions would be redundant with the activation that may already be occurring because of their domain knowledge; thus, the priming tasks should have had no additional effect on their problem solving.

In addition to an induced fixation effect, an incubation effect was observed for the low-knowledge subjects. As predicted, low-knowledge subjects showed more benefit from a delay between problem-solving attempts on both neutral and misleading items. On the other hand, a negative incubation effect was observed for the high-knowledge subjects, as they showed more benefit from the no-delay condition. High-knowledge subjects showed significant improvement on both neutral and misleading items when the second problem-solving attempt immediately followed the first. The fact that a delay had such opposite effects on performance in these two conditions suggests that there may be two quite different ways of escaping fixation. When impasses are generated by particular problem-solving attempts or contexts, they may be more likely to be overcome by incubation. On the other hand, the present results suggest that, when impasses are related to the possession of domain knowledge, a continued or massed effort may allow for more improvement.

MISLEADING EFFECTS ACROSS EXPERIMENTS

A significant misleading effect on the proportion of correct solutions reached by high-knowledge subjects was found in Experiment 1 and replicated in the no-warning condition of Experiment 2, as well as in the nonfixation condition of Experiment 3. Significant effects for the other

dependent variables—proportion of intrusions made and response time—were less consistent. For response time, misleading effects were significant for high-knowledge subjects in Experiment 2 and neared significance in Experiment 3, but not in Experiment 1. High-knowledge subjects tended to make more intrusions on misleading trials, but the effect was not significant in any one experiment. Although effects were not always significant, however, trends consistent with a greater misleading effect for high-knowledge subjects were present. Thus, to test for overall misleading effects for high-knowledge subjects, the data for Experiment 1 and the nontreatment conditions of Experiments 2 and 3 were analyzed together.

A 2×2 (knowledge level \times trial type) ANOVA revealed a significant knowledge level by trial type interaction for all three dependent variables. Across experiments, high-knowledge subjects made fewer correct solutions [$F(1,52) = 17.66$, $MS_e = 0.03$, $p < .0001$], had longer response times on misleading versus neutral items [$F(1,52) = 8.15$, $MS_e = 23.08$, $p < .01$], and were more likely to make intrusions on misleading items than were low-knowledge subjects [$t(52) = 2.25$, $p < .03$]. The overall means for each variable are presented in Table 6.

Additional analyses were performed on just the most knowledgeable subjects. As already noted, the high-knowledge groups in the present experiments scored much lower than the high-knowledge groups from previous baseball-knowledge studies (e.g., Voss et al., 1980). The 10 highest scoring subjects from across the nonmanipulated conditions (Experiment 1, the nonwarning condition of Experiment 2, and the non-fixation-induced condition of Experiment 3) of the three experiments had a mean score on the Baseball Knowledge Questionnaire of 32.3 and were comprised of 9 males and 1 female. This highest scoring group was significantly less likely to solve the misleading problems correctly (.17) than the neutral problems [.59; $F(1,9) = 62.02$, $MS_e = 0.01$, $p < .0001$] and made a significant proportion of intrusions on misleading items [.30; $F(1,9) = 8.75$, $MS_e = 0.05$, $p < .02$]. Further, the highest knowledge subjects took significantly longer on misleading trials (30.5 sec) than on neutral trials [20.7 sec; $F(1,9) = 10.30$, $MS_e = 47.34$, $p < .01$]. The results of these separate analyses confirm that even the highest knowledge subjects experienced significant misleading effects in their problem solving.

GENERAL DISCUSSION

In three experiments, a clear effect of fixation that was due to domain knowledge was observed in problem solving, suggesting that expertise can indeed instill a mental set and promote fixation. Across all experiments, the subjects with the most domain-related knowledge were least able to solve problems correctly when their knowledge suggested an inappropriate solution. Greater domain knowledge was also related to slower overall response times and more intrusions. Even though these additional effects did not reach significance in each individual experiment, when Experiment 1 and the nontreatment con-

ditions in Experiments 2 and 3 were examined together, greater misleading effects were found for high-knowledge subjects than for low-knowledge subjects on all three variables.

The poorer performance of high-knowledge subjects on the misleading problems is consistent with the hypothesis that domain knowledge can act as a mental set. It appears that domain knowledge not only biases a first solution attempt but also fixates the high-knowledge subject by defining and narrowing the search space, preventing a broad search, and decreasing the chances of finding an appropriate solution. Further, in Experiment 2, warning the subjects not to use their knowledge for problem solving did not improve performance on misleading items.

As a result, the subjects at lower levels of knowledge (but not no-knowledge) were more flexible in their problem solving, reaching correct solutions more often than the most knowledgeable subjects. The specific conditions that this effect was observed under were (1) on a creative problem-solving task, where creative is defined as requiring productive (as opposed to reproductive) thinking, and (2) on a task that was of a level of complexity such that novices could engage in problem solving to the same extent as experts.

The inherent tension of prior knowledge as both facilitator and inhibitor of creativity has been of interest since the earliest studies on insight problems (e.g., Maier, 1931, 1945) and has received attention recently within the creative cognition literature (S. M. Smith, Ward, & Finke, 1995; Ward, S. M. Smith, & Vaid, 1997). There is no doubt that extensive domain knowledge is critical for the solution of many problems. Prior knowledge is what provides the solver with possible solutions or associations in the first place. In many cases, more domain knowledge means access to a wider variety of solutions. Within their domains, extensive domain knowledge allows experts to work on problems that would be very difficult or impossible for less knowledgeable individuals to attempt. And, as the complexity of a particular problem-solving task increases, the advantages of extensive domain knowledge in allowing an expert to approach a problem may wash out the potential disadvantages. However, it seems there are also situations where prior knowledge, whether of a domain or of examples, may be detrimental to problem solving, by making the solver vulnerable to fixation effects.

In sum, the contribution of the present study is that it extends earlier work on fixation in problem solving in two ways. First, the present study demonstrated that mental set and fixation may occur as the result of domain knowledge, and second, the third experiment, in particular, suggests that there are important differences between mental blocks that are the result of domain knowledge and those that are the result of the experimental context.

How Is Mental Set Due to Domain Knowledge Broken?

The source of a mental set seems to play an important role in determining how fixation may be best overcome

and creative thinking achieved. Not only was fixation observed in the problem solving of high-knowledge subjects in all experiments, but, in Experiment 3, the fixation manipulation was also successful at fixating the low-knowledge subjects through priming irrelevant solutions. Further, for low-knowledge subjects, the fixation induced by this manipulation was attenuated by introducing a delay between solution attempts. For the high-knowledge subjects, however, fixation was lessened when the second problem-solving attempt immediately followed the first. In other words, low-knowledge subjects demonstrated more flexible thinking as a result of a distributed effort at problem solving, whereas high-knowledge subjects demonstrated more flexible thinking as the result of a massed effort.

One explanation for these findings is that, whereas breaking the novice set may depend on simply waiting for the activation of recently primed irrelevant associations to fade (S. M. Smith & Blankenship, 1991), breaking the expert set may require more active suppression or inhibition of the irrelevant solution paths that are activated by prior knowledge. Since each time an expert newly encounters domain-related information irrelevant solutions may be reactivated by domain knowledge via retrieval from long-term memory, the expert would not benefit from a break between repeated solution attempts, or incubation. Instead, during the course of the problem solving, experts may need to suppress or inhibit the activation for irrelevant solutions that are generated by their knowledge, which would more likely result from continuous or massed solution attempts. Along these lines, Mednick himself (1962) suggested that massed sessions of creative work should be more successful than distributed sessions. Mednick offered the view that it is only once one gets past "the conventional and stereotyped associations to the elements of a problem" that one can begin to entertain the more remote associations which are the key to creative solutions. Especially for experts whose domain-related associations may be quite robust, it may take time to get beyond initial solution attempts, thereby suggesting that experts' creative problem solving would benefit from a massed effort. Thus, the observed negative incubation effect can be seen to be the result of high-knowledge subjects being able to suppress or inhibit the activation of irrelevant solutions only when problem-solving attempts were continuous.

One difficulty with this explanation for the observed negative incubation effect is that the most improvement was seen for high-knowledge subjects on the neutral items. If experts are benefiting from a massed session because it allows them to suppress irrelevant solutions prompted by their domain knowledge, especially performance on the misleading items should improve. It could be that the fixation of the high-knowledge subjects on misleading items may have caused a more general interference on all items and that a prolonged effort allowed the high-knowledge subjects to lessen this overall interference, but an improvement in misleading items along with

neutral items in all conditions would make this a more satisfying explanation.

Instead, the most improvement on misleading items was seen when the high-knowledge subjects were in the fixation condition. In some respects, the fixation condition could be seen as a *hint* where, although the right answer was not given, the wrong answer was, explicitly. Given this hint, high-knowledge subjects in the no-delay condition were the most successful at solving misleading problems. This explanation of how high-knowledge subjects overcome their fixation on misleading problems is consistent with the opportunistic-assimilation hypothesis, which suggests that the advantage of an incubation period is not in providing a time in which inappropriate activations may fade, but in providing an opportunity for the solver to encounter relevant cues for solution (Seifert, Meyer, Davidson, Patalano, & Yaniv, 1995). In support of this explanation, Yaniv, Meyer, and Davidson (1995) found incubation effects on a rare-word definition task only when subjects were exposed to the target word between problem-solving attempts. Outside of highly controlled experimental contexts, it seems quite plausible that the successful solution of previously unsolved problems could depend on a combination of these two mechanisms—that is, both the fading of irrelevant associations and the exposure to relevant cues would improve the chances of successful solution. However, the negative incubation effect observed here among high-knowledge subjects is not fully explained by either the active-suppression or the opportunistic-assimilation hypotheses, or by both of them together. Further investigation is needed to answer the interesting questions of when negative incubation effects might be found in general and why the high-knowledge subjects improved more on neutral items without a delay and without *hints* from the fixation condition in particular.

Expertise as Mental Set

It is important to note that the present study did not investigate the effects of expertise on domain-related creative problem solving *per se*; rather, it provided a demonstration of how the possession of a large amount of domain knowledge may constrain the generation of solutions in the problem solving of experts. There is reason to believe that the mental set that was due to domain knowledge demonstrated in this study may be generalizable across many domains of expertise and problem-solving tasks. The Marchant et al. (1991) result, in which experienced accountants were less able to adapt to a new tax law, can also be seen as an example of fixation, in that experts fail to consider relevant new information because of their previous knowledge. Also, to the extent that experts may be seen to rely heavily on their prior knowledge and less on the specific information given for problem representation, it is plausible that they may be fixated by that knowledge. And there are studies in many domains that suggest that, in fact, experts tend to consider *less* information than do novices in their problem solving. For example, Chiesi, Spilich, and Voss (1979) found that subjects with

high baseball knowledge used less information to make recognition judgements about baseball passages than did low-knowledge subjects. Among similar studies, Devine and Kozlowski (1995) found that subjects with high basketball knowledge made better basketball-related decisions than novices but searched for less information during solution. Experts also used less information than did novices in auditing (Bédard, 1989), medical internship and residency decisions, and financial analysis (Johnson, 1988). Lesgold (1984, personal communication), has suggested that expert radiologists not only spend less time than less skilled individuals looking at X rays and fixate at fewer locations, but also report more information about the X rays than could possibly be available from their fixation points. Isenberg (1986) found that experienced business managers solving simulated management problems used less of the information available and often leapt to solutions before the problems were fully presented. Similar results have been found in relation to medical diagnosis (de Graaff, 1989). It is not just the lack of search but the early commitment to a solution path seen in these last few studies that suggests that experts in a number of domains may be susceptible to mental set and fixation in their problem solving.

Over the past 20 years, many studies have examined how the possession of a large, structured body of domain knowledge enables experts to excel at memory and problem-solving tasks. Previous studies, using the same tests of baseball knowledge as the present study, found that high-knowledge subjects have more extensive, integrated, and conceptual memory of domain-related text (Spilich et al., 1979) and process domain-related text more efficiently (Fincher-Kiefer et al., 1988). However, in the present experiments, the possession of the same domain knowledge led to inferior performance in problem solving. Like other characteristics of expertise, the advantage of extensive knowledge in generating problem representations may benefit problem solving in most domain-related circumstances. Extensive domain knowledge allows experts to infer missing information, make assumptions, and post constraints on a problem space, all of which can lead to a narrower and usually more efficient search for a solution. But the present study suggests that the influence of domain knowledge on generating problem representations may also have its costs, putting experts at a disadvantage when remote associations must be considered or combined in novel ways.

REFERENCES

- ADELSON, B. (1984). When novices surpass experts: The difficulty of a task may increase with expertise. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 483-495.
- ARKES, H. R., & FREEDMAN, M. R. (1984). A demonstration of the costs and benefits of expertise in recognition memory. *Memory & Cognition*, **12**, 84-89.
- BÉDARD, J. (1989). Expertise in auditing: Myth or reality? *Accounting, Organizations & Society*, **14**, 113-131.
- BÉDARD, J., & CHI, M. (1992). Expertise. *Current Directions in Psychological Science*, **1**, 135-139.

- CHASE, W. G., & SIMON, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215-281). New York: Academic Press.
- CHI, M. T. H., GLASER, R., & FARR, M. J. (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.
- CHIESI, H., SPILICH, G., & VOSS, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning & Verbal Behavior*, **18**, 257-273.
- DE GRAAFF, E. (1989). A test of medical problem solving scored by nurses and doctors: The handicap of expertise. *Medical Education*, **23**, 381-388.
- DEVINE, D. J., & KOZLOWSKI, S. (1995). Domain-specific knowledge and task characteristics in decision making. *Organizational Behavior & Human Decision Processes*, **64**, 294-306.
- DUNCKER, K. (1945). On problem solving. *Psychological Monographs*, **58** (Whole No. 270).
- ERICSSON, K. A., & SMITH, J. (1991). *Toward a general theory of expertise*. Cambridge: Cambridge University Press.
- ERICSSON, K. A., & STASZEWSKI, J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 235-268). Hillsdale, NJ: Erlbaum.
- FINCHER-KIEFER, R., POST, T., GREENE, T., & VOSS, J. F. (1988). On the role of prior knowledge and task demands in the processing of text. *Journal of Memory & Language*, **27**, 416-428.
- FRENSCH, P. A., & STERNBERG, R. J. (1989). Expertise and intelligent thinking: When is it worse to know better? In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. V, pp. 157-188). Hillsdale, NJ: Erlbaum.
- GLENBERG, A. M., & EPSTEIN, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition*, **15**, 84-93.
- HECHT, H., & PROFFITT, D. R. (1995). The price of expertise: Effects of experience on the water-level task. *Psychological Science*, **6**, 90-95.
- ISENBERG, D. J. (1986). Thinking and managing: A verbal protocol analysis of managerial problem solving. *Academy of Management Journal*, **29**, 775-788.
- JANSSON, D. G., & SMITH, S. M. (1991). Design fixation. *Design Studies*, **12**, 3-11.
- JOHNSON, E. J. (1988). Expertise and decision under uncertainty. In M. T. H. Chi, R. Glaser, & M. J. Farr. (Eds.), *The nature of expertise* (pp. 209-228). Hillsdale, NJ: Erlbaum.
- LESGOLD, A. (1984). Acquiring expertise. In J. R. Anderson & S. M. Kosslyn (Eds.), *Essays in honor of Gordon Bower* (pp. 31-60). San Francisco: Freeman.
- LUCHINS, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs*, **54** (Whole No. 248).
- MAIER, N. R. F. (1931). Reasoning in humans: II. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology*, **12**, 181-194.
- MAIER, N. R. F. (1945). Reasoning in humans: III. The mechanisms of equivalent stimuli and of reasoning. *Journal of Experimental Psychology*, **35**, 349-360.
- MARCHANT, G., ROBINSON, J., ANDERSON, U., & SCHADEWALD, M. (1991). Analogical transfer and expertise in legal reasoning. *Organizational Behavior & Human Decision Processes*, **48**, 272-290.
- MEDNICK, S. (1962). The associative basis of the creative process. *Psychological Review*, **69**, 200-232.
- MYLES-WORSLEY, M., JOHNSTON, W., & SIMONS, M. A. (1988). The influence of expertise on X-ray image processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 553-557.
- PATEL, V. L., & GROEN, G. J. (1991). The general and specific nature of medical expertise. In K. A. Ericsson & J. Smith. (Eds.), *Toward a general theory of expertise* (pp. 93-125). Cambridge: Cambridge University Press.
- SCHMIDT, H. G., & BOSCHUIZEN, H. P. A. (1993). On the origin of intermediate effects in clinical case recall. *Memory & Cognition*, **21**, 338-351.
- SCHNEIDER, W. (1988). Micro Experimental Laboratory: An integrated system for IBM PC compatibles. *Behavior Research Methods, Instruments, & Computers*, **20**, 206-217.
- SEIFERT, C. M., MEYER, D. E., DAVIDSON, N. S., PATALANO, A. L., & YANIV, I. (1995). Demystification of cognitive insight: Opportunistic assimilation and the prepared-mind perspective. In R. J. Sternberg and J. E. Davidson (Eds.), *The nature of insight* (pp. 65-124). Cambridge, MA: MIT Press.
- SHEPARD, R. N., & METZLER, J. (1971). Mental rotation of three-dimensional objects. *Science*, **171**, 691-703.
- SMITH, S. M., & BLANKENSHIP, S. E. (1991). Incubation and the persistence of fixation in problem solving. *American Journal of Psychology*, **104**, 61-87.
- SMITH, S. M., & SCHUMACHER, J. S. (1992, April). *A test of transfer-appropriate fixation in problem solving*. Paper presented at the 1992 Midwestern Psychological Association, Chicago.
- SMITH, S. M., & TINDELL, D. R. (1997). Memory blocks in word fragment completion caused by involuntary retrieval of orthographically related primes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 1-16.
- SMITH, S. M., WARD, T. B., & FINKE, R. A. (Eds.) (1995). *The creative cognition approach*. Cambridge, MA: MIT Press.
- SMITH, S. M., WARD, T. B., & SCHUMACHER, J. S. (1993). Constraining effects of examples in a creative generation task. *Memory & Cognition*, **21**, 837-845.
- SPILICH G. J., VESONDER, G. T., CHIESI, H. J., & VOSS, J. F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning & Verbal Behavior*, **18**, 275-290.
- STREET, R. F. (1931). *A Gestalt completion task* (Teacher's College Contributions to Education, No. 481). New York: Columbia University, Teachers College Press.
- VOSS, J. F., VESONDER, G., & SPILICH, H. (1980). Text generation and recall by high-knowledge and low-knowledge individuals. *Journal of Verbal Learning & Verbal Behavior*, **19**, 651-667.
- WARD, T. B. (1994). Structured imagination: The role of category structure in exemplar generation. *Cognitive Psychology*, **27**, 1-40.
- WARD, T. B., SMITH, S. M., & VAID, J. (Eds.) (1997). *Creative thought: An investigation of conceptual structures and processes*. Washington, DC: American Psychological Association.
- WEGNER, D. M. (1989). *White bears and other unwanted thoughts*. New York: Viking/Penguin.
- 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.
- WOODWORTH, R., & SCHLOSBERG, H. (1954). *Experimental psychology*. New York: Holt, Rinehart & Winston.
- YANIV, I., MEYER, D. E., & DAVIDSON, N. S. (1995). Dynamic memory processed in retrieving answers to questions: Recall failures, judgments of knowing, and acquisition of information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1509-1521.

APPENDIX
Remote Associates and (Short-Term Memory Task) Stimuli

Word 1	Word 2	Word 3A	Solution A	Word 3B	Solution B
1. LICK	(SPRINKLE)	(MINES)	SALT	(GOLD)	
2. (BAR)	(HARD)	CANE	CANDY	(STOOL)	
3. TYPE	(GHOST)	(STORY)	WRITER	(BOOK)	
4. (CAT)	SLEEP	(BOARD)	WALK	(BLACK)	
5. (SHIP)	(OUTER)	CRAWL	SPACE	(SHAPE)	
6. (THIN)	PICK	(SKATE)	ICE	(ROLLER)	
7. SCOTCH	(WORM)	(RED)	TAPE	(SILK)	
8. (BOTTLE)	TIE	(BREAK)	NECK	(WATER)	
9. (HEARTED)	(FEET)	BITTER	COLD	(BIG)	
10. FAMILY	(APPLE)	(HOUSE)	TREE	(GREEN)	
11. (WILD)	(DARK)	FORK	(PITCH)	SENSE	HORSE
12. (PLATE)	(BROKEN)	REST	(HOME)	SHOT	GLASS
13. (SHORT)	(TRUCK)	SIGN	(STOP)	GUN	HAND
14. (FULL)	(DOWN)	HEAD	(COUNT)	WIND	BLOWN
15. (WORLD)	(CONCERT)	TELEVISION	(SERIES)	TRADE	FREE
16. (FLY)	(BOY)	BEARING	(BALL)	WEIGHT	PAPER
17. (STRIKE)	(WHITE)	LAW	(OUT)	MEDAL	GOLD
18. (CENTER)	(WORK)	CORN	(FIELD)	MEAL	PIECE
19. (FORCE)	(BACK)	ROOM	(PLAY)	CHICKEN	FEED
20. (STOLEN)	(TAX)	ARMY	(BASE)	PRIVATE	PROPERTY

(Manuscript received February 20, 1997;
revision accepted for publication July 11, 1997.)