Facilitating Knowledge Integration and Flexibility: The Effects of Reflection and Exposure to Alternative Models.

Lewis Garnet Roussel Jr
Louisiana State University and Agricultural & Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_disstheses

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_disstheses/6863

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Historical Dissertations and Theses by an authorized administrator of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700  800/521-0600

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
FACILITATING KNOWLEDGE INTEGRATION AND FLEXIBILITY:
THE EFFECTS OF REFLECTION AND
EXPOSURE TO ALTERNATIVE MODELS

A Dissertation
Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Psychology

by

Lewis G. Roussel, Jr.
B.S., University of New Orleans, 1977
M.A., Louisiana State University, 1990
December, 1998
# TABLE OF CONTENTS

LIST OF TABLES ....................................... iv
LIST OF FIGURES ....................................... v
ABSTRACT .............................................. vi
INTRODUCTION .............................................. 1

REVIEW OF THE LITERATURE ................................. 6
   Effects of Rule Search Instructions ..................... 6
   Effects of Providing Explicit Instruction .............. 9
   Effects of Verbalization ................................ 11
   Summary and Implications of Empirical Findings ........ 14
   Summary of Mathews et al. (1996) ........................ 18

EXPERIMENT 1 ......................................... 24
   Introduction ..................................... 24
   Method ........................................... 29
   Results and Discussion ............................. 40

EXPERIMENT 2 ......................................... 55
   Introduction ..................................... 55
   Method ........................................... 56
   Results and Discussion ............................. 58

EXPERIMENT 3 ......................................... 62
   Introduction ..................................... 62
   Method ........................................... 64
   Results and Discussion ............................. 65

GENERAL DISCUSSION ................................... 73
   Summary of Findings ................................ 73
   Explaining Reflective Abstraction Error ............... 83
   Implications ..................................... 93

REFERENCES ........................................... 98

APPENDIX A: COMPUTER SIMULATION PROCEDURES ........... 106
   Translation of Policies into Rules ..................... 106
   Simulation Procedure ................................ 107

APPENDIX B: REPEATED-MEASURES ANALYSIS OF
   PERFORMANCE IN EXPERIMENT 1 ........................ 109

APPENDIX C: ANALYSIS OF FINAL TEST
   PERFORMANCE IN EXPERIMENT 1 ........................ 110

APPENDIX D: ANALYSIS OF FINAL POLICY QUALITY
   IN EXPERIMENT 1 ................................... 111
APPENDIX E: ANALYSIS OF CONTROL GROUPS IN EXPERIMENT 1 ........................................... 112
APPENDIX F: REPEATED-MEASURES ANALYSIS OF PERFORMANCE IN EXPERIMENT 2 .......................... 113
APPENDIX G: ANALYSIS OF FINAL TEST PERFORMANCE IN EXPERIMENT 2 ...................................... 114
APPENDIX H: ANALYSIS OF FINAL POLICY QUALITY IN EXPERIMENT 2 ........................................... 115
APPENDIX I: REPEATED-MEASURES ANALYSIS OF PERFORMANCE IN EXPERIMENT 3 .......................... 116
APPENDIX J: ANALYSIS OF FINAL TEST PERFORMANCE IN EXPERIMENT 3 ...................................... 117
APPENDIX K: ANALYSIS OF FINAL POLICY QUALITY IN EXPERIMENT 3 ........................................... 118
VITA .................................................................................. 119
LIST OF TABLES

1. Mean absolute deviation from target by condition and session in Experiment 1  ...  41

2. Mean absolute deviation from target by condition and session in Experiment 2  ...  59

3. Mean absolute deviation from target by condition and session, collapsed over target level for Experiment 3  ...  66

4. Mean absolute deviation from target by type of hint and target before and after retention interval  ...  70
LIST OF FIGURES

1. General sequence of tasks performed by participants each session ....................... 31

2. Graphical display seen by a participant performing the sugar production task ........... 36
ABSTRACT

Three experiments investigated the effect of various kinds of reflection (within-task and post-task reflection) on learning a process control task. Also, two ways of providing learners with alternative ideas about task behavior, exposure to other's ideas and providing hints for task solution, were examined. The task involved a simulated "sugar production factory" in which the learners sought to control sugar production by manipulating workforce size. It was predicted that combining within-task reflection with exposure to alternative task ideas would lead to superior task performance through integration of experiential and reflective knowledge. Contrary to the prediction, within-task reflection consistently interfered with learning and knowledge integration by causing learners to acquire overly general and invalid rules (reflective knowledge) about the relations among task variables. These results were interpreted as evidence of a fundamental tendency (reflective abstraction error) of people to seek simple relations among variables in complex systems when engaged in within-task reflection. The most efficient learning occurred when reflection was discouraged during task performance, learners were given access to alternative ideas about task behavior, and they were given an opportunity to discuss their task experiences with other learners post-task.
INTRODUCTION

A long-standing view of many investigators of learning is that people acquire and apply knowledge and skills in two distinct and possibly complementary ways (e.g., Anderson, 1982, 1983; Berry & Dienes, 1993, Broadbent, Fitzgerald, & Broadbent, 1986; Brooks, 1978; Cho & Mathews, 1996; Elio & Anderson, 1984; Kolers & Roediger, 1984; Lazarus, 1991; Lewicki, 1986; Mathews, Buss, Stanley, Blanchard-Fields, Cho, & Druhan, 1989; Reber, 1993). One kind of knowledge is direct, not mediated by reflective thought processes such as planning or reasoning. Instead, decisions and actions based on this knowledge are made intuitively, without reflection, and are determined to a great degree by environmental stimuli (Brooks, 1978; Lewicki, 1986; Norman, 1993; Reber, 1989, 1993).

The term used to refer to this kind of knowledge varies depending upon the theoretical position of the researcher. Among the various terms used are procedural (e.g., Anderson, 1976, 1983), nonanalytic (Brooks, 1978), instance-based (Logan, 1988), memory-based (e.g., Mathews et al., 1989), implicit (e.g., Reber 1989, 1993), unselective (Hayes & Broadbent, 1988), and experiential (Norman, 1993). The key attribute of knowledge that is shared by all of these conceptions is that of minimal conscious mediation. That is, knowledge of this type relies heavily on environmental cues to activate the
relevant knowledge; it is highly stimulus-bound. Thus, performance based on this kind of knowledge can be relatively fast and effortless because little reflection is required.

Norman's term, experiential knowledge, will be used herein because it makes the fewest claims about mechanisms underlying knowledge use. A good example of experiential knowledge can be seen in the skill of a chess master scanning the chess board for possible moves. The pattern of chess pieces, as perceived by the player, immediately suggests or affords possible moves. Thus, experiential knowledge influences behavior largely through its effect on perception (e.g., Chase & Simon, 1973; Gibson, 1979; Perruchet & Gallego, 1997).

Experiential knowledge is not taught but is acquired through direct experience (e.g., Gibson, 1979). Furthermore, it is suited best for action, not communication. It may be difficult to express one's experiential knowledge. For example, knowledge engineers involved in constructing expert systems have long struggled with the problem of extracting from experts everything they know about their particular area of expertise (Hoffman, 1987).

Contrast this experiential knowledge with reflective knowledge. Reflective knowledge is suited ideally for thinking and communicating rather than action. When one
performs a task using reflective knowledge (e.g., following a recipe), the performance is mediated by reflection, as opposed to the use of experiential knowledge which is immediate (Anderson, 1982, 1983; Norman, 1993).

By definition, reflective knowledge is accessible to consciousness. In fact, in the scientific community it is the primary means of communication. The theory is the quintessential example of reflective knowledge. Reflective knowledge is suited ideally for an existence in the public space where ideas are developed and refined.

The very quality which gives reflective knowledge its power, its abstractness, can also be its weakness. Reflective knowledge may be misapplied because the conditions of its use may have been abstracted away. It is incomplete because it is abstract and therefore missing many possibly important details (Ornstein & Ehrlich, 1989). Even though the abstract quality of reflective knowledge should enable transfer across domains, it's very abstractness may actually prevent it from being accessed in relevant situations. The literature on analogical transfer demonstrates the difficulty of transferring reflective knowledge (e.g., Catrambone & Holyoak, 1989; Gick & Holyoak, 1980). Typically, one must be told of the relevance of one's knowledge for the transfer problem. Otherwise, no transfer occurs.
The kind of thinking required of reflective knowledge may demand too much of people’s limited capacity memory and attention (Hayes & Broadbent, 1988; Miller, 1956). The modern development of cognitive artifacts ranging from paper and pencil to computer applications used for decision support is testimony to the difficulty of reflective thinking as well as its importance.

The complementary nature of experiential and reflective knowledge suggests that integrating the two would be beneficial for learning and transfer. For example, the context-bound character of experiential knowledge could help ground reflective knowledge and prevent its misapplication or over-generalization. Conversely, reflective knowledge could facilitate the transfer of experiential knowledge into novel but related situations. These positive outcomes of combining experiential and reflective knowledge have not been born out by studies, however. As a subsequent literature review will show, successful integration of the two kinds of knowledge is not easily achieved. Learners tend to default to experiential modes of thinking when reflective strategies fall into difficulty (Berry & Dienes, 1993; Reber, 1993). Also, when learning to perform complex tasks, learners may acquire invalid reflective knowledge in the form of mental models or verbalizable rules that lead to less than optimal performance (Reber, 1976; Reber, Kassin, Lewis, & Cantor, 1993).
1980). Finally, dissociations between experiential and reflective knowledge may occur, particularly in situations where the rules governing task behavior are not salient (Berry & Broadbent, 1984; Nisbett & Wilson, 1977).

Understanding how and when integration fails or succeeds can lead to improvements in methods of training and educating and increase our basic understanding of learning and transfer. The intent of this study was to investigate how learning and transfer is affected by encouraging various kinds of within-task and post-task reflection and by giving learners access to alternative ideas about how to perform a task.
REVIEW OF THE LITERATURE

This literature review focuses on the effects of explicit or reflective processes on the resulting knowledge, performance, and transfer for tasks in which dissociations (i.e., lack of integration) frequently have been reported. The studies to be discussed used three methods for eliciting reflective thinking: (1) instructing participants to discover the rules or structure that determines the behavior of the task variables; (2) providing explicit instruction or information about the rules that relate task variables; and (3) requiring participants to verbalize as they perform the task.

Following the discussion of the literature, the implications of these findings for the integration of knowledge will be discussed. Finally, the results of three experiments will be presented.

Effects of Rule Search Instructions

One way to investigate the influence of reflective thought on experiential learning is to instruct participants to attempt to figure out the rules governing the behavior of the task. The results of this approach have been mixed. The effect of rule search instructions has ranged from decreasing the level of learning (e.g., Berry & Broadbent, 1988; Howard & Ballas, 1980; Reber, 1976; Reber et al., 1980) to having no effect (Dienes, Broadbent, & Berry, 1991; Dulany, Carlson, & Dewey, 1984).
or improving learning (Berry & Broadbent, 1988; Reber et al., 1980). The particular effect those rule-search instructions produce seems to be related to the salience of the rules governing the task (Berry & Broadbent, 1988; Mathews et al., 1989; Reber et al., 1980).

Reber et al. (1980) investigated the effect of salience on people's ability to learn an artificial grammar. Reber and his colleagues presented half of their participants with training stimuli arranged randomly in an array to make less salient the common features (i.e., patterns of letters in grammatical exemplars) that could be used to facilitate rule learning. Stimuli for the remaining participants were ordered in a way that maximized the salience of common features. Learning was measured by requiring participants to classify new stimuli as valid or invalid according to the rules. These authors reported that rule search instructions reduced learning in the random presentation condition only. A more fine-grained analysis of the data showed that participants in the random presentation conditions tended repeatedly to misclassify stimuli that were presented more than once during test. This suggested that the participants had acquired invalid rules which they were using to incorrectly classify certain stimuli. Thus, the likelihood of acquiring invalid knowledge can be increased if stimulus structure is difficult to detect and explicit processing, such as rule-search, is invoked.
When the rules governing relations among the stimuli are salient or easy to discover, rule-search instructions can have a positive effect on learning (Mathews et al., 1989; Lee, 1995; Reber et al., 1980). Mathews et al. (1989) used a biconditional artificial grammar built from a set of correspondence rules described below. This grammar generated letter-strings consisting of eight consonants, with a dot separating the first and last four letters (e.g., "SCCT.VPPX"). There were no constraints on the four letters to the left of the dot, except that they had to be from the set of letters reserved for the grammar. The correspondence rules specified which letters must occur in corresponding positions in the left and right halves of the string. For example, $S$ goes with $V$, $C$ with $P$, and so on. Thus, in the example above, an $S$ in the first position requires a $V$ in the corresponding position in the right half of the string.

One likely effect of the correspondence rules was that salient patterns such as repeated letters, "SCCC.VPPP" for example, focused the learner on the relationships across the left and right halves of the strings. This made it more likely that learners would discover the rules. Two training tasks were used to elicit either an explicit rule-discovery strategy or an implicit exemplar memorization strategy. Learning by participants in the rule-discovery condition was superior to that of participants in the
exemplar memorization conditions. This finding is consistent with the claim of Reber et al. (1980) that rule-search would be successful only when the rules are easy to discover.

Rule-search instructions do not always affect performance in implicit learning tasks (Dulany et al., 1984; Lee, 1995; Mathews et al., 1989). In learning tasks involving rules that are extremely difficult to find, participants are likely to fall back on an implicit or experiential mode to guide their performance (e.g., Berry & Dienes, 1993; Reber, 1993). The deliberate engagement in reflective or explicit thinking is effortful and less fun than experientially performing the task (Ericsson, Krampe, & Tesch-Romer, 1993) or being in flow (Csikszentmihalyi, 1990). Learners know that progress can often be made by mere repetition of the task, without reflective thinking. Thus, there is a strong natural tendency to perform a task without attempting effortful reflective strategies.

Effects of Providing Explicit Instruction

A few studies have provided specific information to participants about the nature of the rules underlying the task (Berry & Broadbent, 1984; Reber et al., 1980; Stanley, Mathews, Buss, & Kotler-Cope, 1989). Berry & Broadbent (1984) used the sugar production task in which participants attempted to reach a specified target sugar production level in a computer simulated sugar refinery by
manipulating the size of the workforce employed at the refinery. Sugar production and workforce size were related by a simple mathematical equation. On each of a series of trials, the computer displayed the current sugar production. The participants were required to input the size of a work force they thought necessary to reach the target production level. Then they received immediate feedback about the actual production level achieved.

One aim of Berry and Broadbent's (1984) study was to find out if verbal instruction on how to reach the target would affect task performance and verbalizable knowledge similarly. They found that verbal instruction improved their participants' ability to answer questions about task behavior but it had no effect on their ability to control sugar production, except when combined with a requirement to verbally justify each response.

Stanley et al., (1989), also used the sugar production task to investigate the effect of several kinds of instruction such as exemplar memorization, providing a simple heuristic, rule instruction, and providing written transcripts gathered from subjects "thinking aloud" as they performed the task. Stanley et al., (1989) found a small but statistically significant benefit from all of these instruction types relative to a control condition. Interestingly, performance in rule instruction conditions was no better than in exemplar memorization conditions even
though it provided the participants with reflective knowledge about relations among task variables. This is consistent with the hypothesis of separate kinds of knowledge. Verbal instruction primarily influences reflective knowledge whereas task performance is based primarily on knowledge gained through task experience (experiential knowledge). Sometimes, even perfect reflective knowledge of task behavior can be useless without real task experience (Reber & Millward, 1968).

**Effects of Verbalization**

Another way to encourage participants to think and reflect as they learn is to require them to think out loud and report their reasoning as they perform the task. It has been suggested that the requirement to verbalize while learning a task might improve learning by maintaining the learner's focus on the relevant variables, assuming the relevant variables are salient or likely to be selected by the learner (Berry & Broadbent, 1984). Berry and Broadbent (1984) reported that concurrent verbalization was not effective in improving learning in a complex control task (e.g., sugar production task). However, their experiments lacked a proper nonverbalizing control. Subsequent investigations, using similar tasks, showed verbalization to cause a small but positive increase in the level of learning (McGeorge & Burton, 1989; Stanley et al., 1989).
It is difficult to interpret the results of these studies because the methods of eliciting verbalization differed widely across studies. Berry and Broadbent's (1984) participants were asked to give a reason for each response as they interacted with the task. McGeorge and Burton (1989) asked their participants to describe any rules or heuristics they were using. In both of these studies participants verbalized concurrently with task performance. By contrast, the participants in the experiments of Stanley et al. (1989) verbalized instructions for an unseen partner to perform the task after each ten interactions with the task.

That these different methods all led to improved performance (except Berry & Broadbent) suggests that any kind of verbalization is sufficient for improvement in performance. However, some findings from problem solving research suggest that some forms of verbalization are better than others (Berardi-Colletta, Buyer, Dominowski, & Rellinger, 1995). Berardi-Coletta et al. (1995) hypothesized that problem solving would be enhanced only when participants engaged in "metacognitive processing." They described metacognitive processing as a reflective process of attending to what one is doing to solve the problem (process-focused) rather than focusing on aspects of the problem (problem-focused). Berardi-Colletta et al. (1995) elicited process-focused behavior by periodically...
prompting their participants to justify and evaluate their actions. Problem-focused behavior was elicited by having participants periodically report on various aspects of the problem (e.g., the goal or the current problem state). In support of their hypothesis, Berardi-Colletta et al. (1995) found that problem solving performance was superior for participants in process-focused conditions, relative to problem-focused conditions. Berardi-Colletta et al. (1995) also reported that process-focused participants were more likely to change their strategies and problem representations when these were found to be inadequate for problem solution. By contrast, problem-focused participants persisted in inefficient strategies and problem representations.

Other investigations have found problem solving performance to be positively related to engagement in self-dialogue or generating explanations about task behavior (Biemiller & Meichenbaum, 1992; Chi, Bassock, Lewis, Reimann, & Glaser, 1987; Chi, DeLeeuw, Chiu, & LaVarcher, 1994). For example, Chi et al. (1987) collected think-aloud protocols of students studying example problems. They found that better problem solvers generated more statements indicative of self-monitoring and self-explanation, absolutely and proportionally, than poor problem solvers. Also, the number of self-monitoring and self-explanation statements correlated highly with problem
solving performance. More recently, Chi et al. (1994) reported that participants prompted for self-explanations showed greater understanding of studied materials than participants not so prompted.

Summary and Implications of Empirical Findings

There are several key findings that bear on the problem of integrating experiential and reflective knowledge. First, reliance on experiential modes of thinking and experiential knowledge is a default strategy when reflective strategies run into difficulty (Reber, 1993). Second, when the task is structured so that the rules and relevant variables are not salient and quality feedback is missing, learners often end up with invalid reflective knowledge or mental models (Hayes & Broadbent, 1988; Reber et al., 1980). Further, these invalid models are often resistant to challenge or change (Berardi-Colletta et al., 1995; Holland, Holyoak, Nisbett, & Thagard, 1986). Third, providing reflective knowledge, in the form of instruction, tends to affect the learner's mental model and leave experiential knowledge unaffected (Berry & Broadbent, 1984). Last, thinking about one's knowledge (i.e., process-focused) in relation to evidence (experiential knowledge) rather than thinking with it (e.g., problem-focused) is most likely lead to revision of one's model (Berardi-Colletta et al., 1995; Kuhn, 1989).
These findings suggest three conditions necessary for integration of reflective and experiential knowledge: (1) Provide support for reflective thinking so learners will not default to experiential mode, (2) provide exposure to alternative task conceptions (mental models) along with accurate and timely feedback for evaluating models, and (3) provide an environment for learners to think about their models in relation to the task as they practice.

Support for Reflective Thinking

Relying on experiential knowledge requires fewer resources than reflecting on and reasoning from a mental model of the task. Performing a task in experiential mode requires only attending to the task; the task itself provides the retrieval cues for past solutions. In contrast, performing in reflective mode requires maintaining in conscious memory a model of the task (see Norman, 1993 for a persuasive argument about the extra effort required for reflection). Thus, some kind of support for reflection might lessen the tendency to fall back on experiential strategies for learning. However, merely instructing participants to learn reflectively will likely fail. Therefore, the reflection must be "built in" to the learning task.

Exposure to Alternative Models

When learners do seek rules governing task behavior or otherwise build a mental model of the task, and there is a
lack of timely and accurate feedback, the resulting rules or model can be flawed. Task experience is not sufficient to cause revision of the learner's model for a number of reasons. In some cases there may be insufficient or unreliable data for modifying the model (Brehmer, 1980). At times, repeated task experience may entrench one's model and block solution of problems easily solved otherwise (e.g., Luchins, 1942). At other times, one's model may resist repeated opportunities for change by creating "exception rules" (Holland et al., 1986). Thus, an increasing level of experience is not necessarily associated with an increasing correctness of one's model.

Regardless of the quality of the data confronting their mental models, people most often are biased toward evidence that confirms their model (e.g., Wason, 1968). Therefore, the learner is unlikely to revise a faulty model or to consider alternatives (cf., Kareev & Avrahami, 1995).

Learners also may lack reflective skills. Rather than reflecting about their models and the relevant evidence, individuals lacking reflective skills might reflect with their models (Kuhn, 1989). Thus, the model is never a candidate for modification as it is largely invisible to the thinker.

Overcoming these obstacles to successful model revision is necessary for integrating reflective and experiential knowledge. To achieve this the learner must be encouraged
to consider alternative models and must have available accurate and timely feedback about the quality of each model.

**Thinking About Instead of Thinking With**

There is no guarantee however that providing alternative models along with accurate feedback will have the desired result. It is possible that providing this extra support to learners will affect reflective knowledge only, leaving experiential knowledge, and therefore task performance, untouched. This might be prevented by providing a means for learners to use their reflective knowledge, and reflect on that knowledge, as they practice the task (acquire experiential knowledge).

A recent investigation of Mathews et al. (1996), using the sugar production task, tested a procedure devised to meet the three conditions just discussed. The procedures used by Mathews et al. (1996) included group discussions among learners to formulate explicit models of task performance. The participants were required to express their explicit models in the form of a written policy for performing the task. The group discussions were expected to provide participants with an opportunity for reflection about their models and to make available alternative models (from other group members). Additionally, Mathews et al. (1996) gave feedback to learners about the quality of their
explicit models thereby raising the quality of evidence available to learners for evaluating their models.

Mathews et al. (1996) also incorporated a reflective practice procedure designed to facilitate integration of the learners' mental models with their experiential knowledge. This was predicted to reduce the dissociation between reflective and experiential knowledge. These procedures produced substantial improvement in performance when the participants received all of these treatments. The current investigation attempted to replicate and extend the work of Mathews et al. (1996). Therefore, the findings of Mathews et al. (1996) will be discussed in some detail below.

Summary of Mathews et al. (1996)

The task used by Mathews et al. (1996) is a dynamic systems control task in which participants learn to control sugar output in a computer simulated sugar production factory (Berry & Broadbent, 1984, 1988). Briefly, participants interact with a computer-based simulation of a sugar production factory. The task requires participants to control the level of sugar production by setting the level of work force. Sugar production is computed as a function of sugar production on the previous trial and work force on the current trial. The equation relating sugar production to work force is $P_1 = (20 \times W) \cdot P_0 + N$. $P_1$ is the
sugar production output ranging from 1000 to 12000. \( P_0 \) is the sugar production level at the beginning of the trial and \( W \) is the number of workers entered by the participant. \( W \) ranges from 100 to 1200. \( N \) is a random element that is added to the sugar production output. This value is randomly selected on each trial from the values 1000, 0, and -1000. The random element functions to make the relation between production and workforce less salient.

Sugar production is not allowed to exceed 12000 tons nor go below 1000 tons. In those cases where a participant's response results in sugar production outside of this range, the production output is truncated to the appropriate maximum or minimum value. Each time the participant chooses and enters a value for workforce, the computer calculates the new sugar production level and displays it on the computer screen.

Mathews et al. (1996) introduced a novel reflective practice procedure to force participants to think about the relation between their reflective knowledge (as represented in their written strategies or policies) and how they actually performed the task. During the interval between practice sessions, Mathews et al. evaluated each policy via computer simulation. The results of the simulations were then used to provide feedback to the participants about the quality and behavior of their policies. On each trial,
participants were required to predict how many workers their written policy would use (as determined by the simulation), and what particular rule of their policy would be employed to set worker level on that trial. Then, they were given an opportunity to suggest a different worker level that might get them closer to the target level of output. Following their responses, they received feedback on what level of workers their policy would have prescribed in this situation (as determined by the computer simulation), which rule or statement of their policy would have been used to set worker level, and the outcomes in sugar production levels for both their policy’s response and their alternative response (if one was suggested).

Two conditions were compared. In one condition (singles) participants worked alone. These participants formulated their own individual policy for controlling sugar production after each of the three practice sessions. The practice sessions were distributed over a one-week period with one session every other day. Participants in the groups condition worked together in teams of four or five individuals after each practice session to create a group policy. Both singles and groups used the reflective practice procedure in which they predicted what their previous policy would do during the second half-hour of the second and third practice sessions (they had no policy to reflect on in the first practice session).
Results

The policies generated by participants in the group's condition were superior (better at controlling sugar production as determined by computer simulation) to those in the singles condition. Similarly, mean individual performance for participants in the groups condition was better than that in the singles condition.

Mathews et al. (1996) used a predetermined cutoff score to quantify the number of expert performers and policies in each condition in the final session of practice. Fifty-nine percent of the participants in the groups condition were classified as expert performers compared to 13 percent in the singles condition. The difference was statistically reliable. All of the groups achieved expert level policies whereas none of the singles' policies achieved expert level.

Discussion of Results of Mathews et al. (1996)

Apparently, group interaction was essential for obtaining the best results in terms of the most expert level policies and for improving individual performance. The group interaction presumably was effective because it helped overcome rigidity in individual approaches to the task (i.e., negative effects of mental sets, e.g., Norman, 1993; Luchins, 1942) and it exposed learners to new and different ideas for performing the task. However, it must be noted that the group interaction variable was confounded.
with exposure to alternative models. That is, participants in the *singles* condition did not have access to any policies other than their own. It is possible that exposing participants in the *singles* condition to the ideas of other participants in the absence of group interaction (e.g., give each participant a written text of other policies) would have produced benefits as well.

The findings of Mathews et al. also suggest that reflective practice was essential for improvement in policy quality and task performance. Mathews et al. argued that reflective practice helped learners integrate new ideas (reflective knowledge) with experiential knowledge that drives task performance. As suggestive as their results are, there was no direct comparison of reflective practice to a control condition in which no reflective practice was done.

Therefore, one of the major goals of this investigation was to attempt to replicate the findings of Mathews et al. (1996) while adding an appropriate nonreflective control condition and removing the confound of the group interaction variable and exposure to alternative models. That goal was addressed in Experiment 1.

The results of Experiment 1 failed to replicate the findings of Mathews et al. (1996). The results suggested that the lack of effectiveness of reflective practice might
be related to the kinds of models learners use when reflecting. This was explored in Experiments 2 and 3.

As the general intent of this investigation was to examine how learning and transfer are affected by reflection and access to alternative ideas, the third experiment included a transfer task at the end of training.
EXPERIMENT 1

Introduction

Experiment 1 attempted to replicate and extend the findings of Mathews et al. (1996) in which a combination of reflective practice, exposing learners to alternative models, and group discussion led to superior performance relative to a condition in which learners were exposed only to their own ideas. In particular, it investigated the effect of reflection (i.e., writing down one's strategies and discussing ideas with group members), exposure to alternative models, and reflective practice on performance in the sugar production task.

Reflection

One kind of reflection occurs in the process of writing down one's ideas. Requiring participants to produce written statements of their task knowledge can be used as a way of assuring that reflection occurs. At the end of each session in this experiment, participants in all but two control conditions were required to write down a set of instructions, a "policy", for performing the task. Thus, each policy was taken as evidence for reflective behavior and was assumed to provide information about the nature of each participant's mental model of the task.

It was predicted that this reflection would be insufficient for improving task performance primarily because it occurred post-task; policies were written at the
end of each practice session. Gagne and White (1978) suggested that successful instruction occurs only when propositional (reflective) knowledge and intellectual skills (i.e., experiential knowledge) are linked together through reflection concurrent with practice.

Exposure to Alternative Models

Exposure to alternative models may be necessary to counteract people's tendency to seek only information that is consistent with their view (Wason, 1968). To provide the opportunity for participants to encounter alternative task solutions and reflect about them, Mathews et al. (1996) had their participants meet periodically in small discussion groups. The goal of these discussions was to produce a written policy describing how best to achieve target production. Mathews et al. (1996) reported much better performance in a condition involving group discussion relative to a condition in which learners worked individually. However, it is not clear whether the superior performance was due to the discussions or to the exposure to others' ideas that occurred in the discussions. Therefore, in Experiment 1 a condition was included in which participants were given copies of others' policies but not allowed to discuss those policies with other participants. This condition was called nominal groups to indicate that they had benefit of the group's ideas without actually meeting as a group. The nominal groups condition
was compared to a group interaction condition in which discussion took place. Thus, the participants in the group interaction and nominal groups conditions were all exposed to alternative models but group discussions took place only in the group interaction conditions.

Exposure to alternative models was predicted to improve the quality of individual policies. This prediction follows from finding that people will seek disconfirming evidence of their hypotheses if they are made aware of the existence of alternative hypotheses (Kareev & Avrahami, 1995). Thus, policies should improve if invalid rules (hypotheses) contained in them can be rejected in favor of better rules in other's policies. The degree to which exposure to alternative models facilitated performance (experiential knowledge) was expected to depend on the presence of a third factor, reflective practice. Only when reflection and practice occur together is it expected that the two sources of knowledge become integrated (Gagne & White, 1978).

**Reflective Practice**

Reflective practice was defined to be performance of the task under conditions in which task decisions are based on one's mental model (reflective knowledge) and one attends to the consequences of those decisions. Berardi-Coletta et al. (1995) found that requiring their participants to "think aloud" was not effective for
increasing transfer performance in a problem solving task. However, when their participants were required to explain their thinking, thereby invoking metacognitive processes such as planning and monitoring, transfer was significantly improved. In Experiment 1, the effect of reflective practice was assessed by comparing the performance of participants who used a reflective practice procedure for training versus participants practicing the normal sugar production task (experiential practice).

One potential drawback of the reflective practice procedure of Mathews et al. (1996) was that participants were not free to choose which policy statement to evaluate. Instead, they attempted to predict which policy statement would be applied by a computer simulation of their policy. In effect, the participants in Mathews et al. were attempting to guess what an expert, exemplified by the program simulating their policy, would do in the current situation if restricted to using only the participant's policy. Berry and Broadbent (1987) found that participant control of the timing of explanations, as opposed to experimenter control, was important for learning. Also, according to some recent theorizing in Artificial Intelligence (AI), users of expert systems should be more likely to accept and therefore use a system if they are allowed to participate in the construction of explanations of the behavior of the expert system (Ford, Canas, &
Coffey, 1993). Possibly, giving the participants control over which policy statement to evaluate would help them remember the status of their policy statements as they worked. Therefore, in the current experiment, the reflective practice procedure of Mathews et al. (1996) was modified by requiring the participants to choose a policy statement for evaluation. Subsequently, they had to predict the outcome of applying the policy statement. In order to distinguish between the procedure used by Mathews et al. (1996) and the modified procedure used in this experiment, the term "assisted reflective practice" will be used to refer to the modified procedure.

In summary, the first experiment investigated the roles of reflection, reflective practice, and exposure to alternative models when performing a complex control task (e.g., the sugar production task). Three predictions were made. First, improvements in ability to control sugar production would occur only when exposure to alternative models was accompanied by reflective practice. Similarly, the mental models of participants (i.e., written policies) were expected to be influenced only by the combination of reflective practice and exposure to alternative models. Third, combining reflective practice with exposure to alternative models was expected to lead to greater integration of reflective and experiential knowledge.
Method

Participants and Design

One hundred fifty-three undergraduate students enrolled in introductory psychology courses at Louisiana State University were recruited to voluntarily participate in return for extra-credit.

The experiment was arranged as a factorial design comprising three factors: exposure to other's policies (group interaction vs nominal groups vs no exposure), practice mode (assisted reflective practice vs experiential practice), and session. Additionally, two control groups, practice-only and predict-only, were included. Neither control condition required participants to write a policy. The practice-only condition is a baseline against which to compare the effect of reflective practice.

The two control conditions differed from one another only in the practice task performed each session. During practice, participants in the predict-only condition were required to predict the resulting sugar production on each trial of practice (prediction task). Comparisons between the practice-only control and the predict-only control were used to measure the effect of unassisted reflection. Comparisons between the practice-only control and the no exposure/experiential condition were used to evaluate the effect of writing a policy.
The two primary dependent variables were performance as indicated by the average unsigned deviation from target production during the test phase and quality of the final policy. Policy quality was measured by using the policy to simulate performance of the sugar production task. The average unsigned deviation from target production achieved by the simulated policy was taken to be the policy quality. The simulation procedure is described in Appendix A.

Procedure

Participants were tested in groups ranging from three to six individuals. Each group was randomly assigned to one of the eight conditions. Regardless of condition, all participants completed three sessions, one per day. For all participants, the three sessions were completed within seven days. Figure 1 depicts the sequence of tasks performed in each session by the participants. As shown in Figure 1, participants performed three basic tasks each session: (1) practice; (2) test; and (3) writing individual policies. Additionally, participants in the nominal groups and group interaction conditions began the second and third sessions by forming a group policy.

First session. In the first session, all participants were told that they were to take on the role of manager of a simulated sugar production factory. They were informed that their job was to learn how to achieve and maintain a target level of sugar production by interacting with the
Figure 1. General sequence of tasks performed by participants each session in Experiment 1. Participants in the no exposure conditions did not construct group policies.
simulation. They were further informed that the only variable they could control was the size of workforce used at the simulated sugar production factory. Thus, their task was to learn the relationship between workforce size and production level. Participants were also told that they would have additional tasks but would be given instructions at the appropriate time.

After receiving instructions, participants were given 30 minutes to practice or interact with the simulation program. In all conditions but predict-only the simulation program was the standard sugar production task. In the predict-only condition, the prediction task replaced the sugar production task as the practice task. The sugar production task and prediction task are described below.

After practicing for 30 minutes, all participants performed the test. The test comprised four blocks of 10 trials of the sugar production task. The participants were allowed up to 30 minutes to complete the test. The participants were informed that their goal was to stay as close as possible to the target production level.

After completing the test, each participants, except those in the two control conditions, were given 15 minutes to write down his or her policy for controlling sugar production. This completed the first session.

Second and third session. In all conditions but the two control conditions, participants were returned their
written policies from the previous session. The policies were inscribed with a rating. Participants were told that the rating indicated how well another manager was able to perform the sugar production task when using their written policy as a guide. The rating procedure is described below under the heading of policy feedback.

After receiving an explanation of the ratings, participants in the nominal groups and group interaction conditions were given 15 minutes to form a group policy before proceeding to the practice-test-write policy sequence followed in the first session. The group policy formation procedure is described below. Participants in the no exposure conditions did not construct group policies.

The same practice-test-write policy sequence used in the first session was followed for the second and third sessions. However, before beginning to practice, all participants (except controls) were told that they would have to write a new policy at the end of the session and therefore should be thinking about how to improve their policy as they practiced.

To investigate the effect of reflective practice, the practice task was changed as follows for the second and third sessions. Participants in reflective practice conditions (except for controls) performed the assisted
reflective procedure instead of the sugar production task
during the practice portion of the session.

The test portion of the second and third sessions was
the same 40 trial sequence as in the first session.
However, this time, participants were allowed to refer to
their written policies from the previous session as they
performed the test.

At the end of the session, participants were instructed
to write new policies based on the performance of their old
policies. They were informed that they could include any
or all of their old policies in the new ones. After the
end of the third session, all participants were debriefed
and given a slip for their extra credit points.

Tasks

Sugar production task. Experiment 1 used the sugar
production task used by Berry and Broadbent (1984) and as
described in the discussion of Mathews et al. (1996). This
is a computer-based task in which participants imagine they
are in charge of a factory that produces sugar. The
participants attempted to achieve and maintain a specified
level of an output variable, sugar production, by
controlling the number of workers employed by a simulated
sugar-production factory. On each task trial, the computer
presented to the participant a current level of sugar
production. The participant responded by choosing and
entering the number of workers to be employed. The
computer then updated and displayed the new sugar
production level.

Task trials were grouped into blocks of ten trials and
each block began with a randomly selected production level.
Figure 2 shows the graphical display seen by participants.
Sugar production level is represented on the vertical axis
of the graph. The dashed horizontal line shows the target
production level. The horizontal axis represents the
sequence of trials. The number of workers entered on each
trial is displayed on the horizontal axis. Each sugar
production output is represented by an 'X' on the graph.
At the end of each block, the display was cleared and a new
graph displayed for the next block of trials.
Sugar production was allowed to vary from 1000 tons to
12000 tons. Participants were allowed to select a number
of workers ranging from 100 to 1200 in multiples of 50.
The target production was fixed at 6000 tons.

The relationship between number of workers and sugar
production was identical to that used by Mathews et al.
(1996). The main dependent measure was the mean unsigned
deviation from target production, in tons, across a block
of ten trials. Because the target production level was
always 6000 tons, the dependent measure could vary from a
minimum of zero, if on target for every trial, to a maximum
of 6000. Chance performance was defined as the mean
unsigned deviation that would be achieved by entering a
Current production level is 2000 tons.  
Current workforce size is 300 workers.

Enter number of workers = » _

Figure 2. Graphical display seen by a participant performing the sugar production task. The target production is represented by the dashed horizontal line and the production level achieved on each trial is indicated by an 'X'. Number of workers used on each trial is recorded below the horizontal axis.
random value for workers on every trial. Chance performance was thus determined to be 4206 tons.

In order to minimize variability in the dependent measure, the effect of the random element was removed from the deviation score before submitting the data for analysis. The random element was removed by recomputing the production level that would have been achieved on each trial without the random element in the equation.

**Assisted reflective practice.** The assisted reflective practice procedure required participants to explicitly state which particular statement of their written policy (their reflective knowledge of the task) they were following when they made their decisions about the number of workers to use. On each trial, the computer prompted them to enter an identifier (statement number) for the statement or rule they thought best applied in the current situation. They were also prompted to enter the number of workers to be used according to the rule, and the production level they expected to achieve. After entering this information, the computer calculated and displayed the new production level. At this time, the participant was given the opportunity to repeat the trial by entering a different number of workers.

**Prediction Task.** Participants in the predict-only control condition performed a modified version of the sugar production task during their practice sessions. On each
trial of the prediction task, after entering a worker level, they were prompted by the computer to predict what the resulting production would be. After entering their prediction, the computer calculated and displayed the new production level.

**Policy feedback.** During the one-day interval between sessions, the written policies were assigned a rating indicating their quality. At the beginning of all but the first session, the rated policies were returned to the participants in order to give them feedback about the validity of their reflective knowledge.

The ratings were determined by using them to perform the sugar production task for 20 trials. On each trial, a rater selected the most appropriate rule from the policy and entered the indicated number of workers. The most appropriate rule was considered to be the one that matched the current situation and was the most specific in its range of application. Consider, for example, the two following rules: (1) "if you are above the target production of 6000 then you should decrease the size of the workforce"; and (2) "if current production level is between 8000 and 10000 tons then you should use 800 workers." Both rules would be applicable to any trial on which current production level is 9000 tons. However, the second rule is more specific (i.e., applicable in fewer situations) and would be chosen by the rater. On trials where no rule
applied, the rater entered the same number of workers used on the previous trial, unless it was the first trial. In this situation the rater entered a randomly selected number of workers. On trials where the policy indicated only a range of workers (e.g., more workers, or a high number of workers) the following actions were taken: (a) "more, or less, workers than X" was interpreted as a randomly selected value of workers between X and the maximum or minimum number of workers allowed, respectively; (b) "a high, or low, number of workers" was taken to mean a randomly selected number of workers above 750 or below 450 respectively; and (c) "an increasing, or decreasing, number of workers" was interpreted the same as in (a). A random number generator (computer program) assisted the rater in selecting random values.

Each policy was assigned a rating indicating its quality according to the average unsigned deviation from target. The policy ratings, described below, were inscribed on the policies. Policies were rated as follows: **EXPERT** - deviation less than 1000 tons, **GOOD** - deviation greater than or equal to 1000 tons and less than 2500 tons, **SATISFACTORY** - deviation greater than or equal to 2500 tons and less than 4000 tons, and **UNSATISFACTORY** - deviation greater than or equal to 4000 tons.

**Forming group policies.** At the beginning of the second and third sessions, and after receiving copies of their
rated policies written at the end of the previous session (see Figure 1), participants in the group interaction conditions met for 20 minutes in order to form a group policy. Interactions of the groups were unstructured, that is, members were free to organize the process in any way they chose. The experimenter instructed them to combine their individual policies into a single written policy. After completing this process, the experimenter made copies of the group policy and distributed one to each group member to use during the subsequent practice and test procedures.

Participants in the nominal groups conditions did not discuss their policies. Instead, they received copies of the individual policies of the other members of their group and were instructed to form a single written policy based on those policies. They were instructed to use any rules from their own or others' policies that they wished.

Results and Discussion

The Effect of Assisted Reflective Practice on Reflective Knowledge

The final policies for all participants were evaluated by a computer simulation as described in Appendix A. The mean simulated performance of the final session policies is displayed in Table 1. Data from the two control conditions were not included in the analysis of reflective knowledge because participants did not write policies.
Table 1. Mean absolute deviation from target by condition and session in Experiment 1.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Session</th>
<th>Final Policy</th>
<th>( r )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>no exposure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>experiential</td>
<td>28.0</td>
<td>21.7</td>
<td>16.9</td>
</tr>
<tr>
<td>reflective</td>
<td>26.2</td>
<td>26.2</td>
<td>23.9</td>
</tr>
<tr>
<td>nominal groups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>experiential</td>
<td>27.9</td>
<td>20.8</td>
<td>19.7</td>
</tr>
<tr>
<td>reflective</td>
<td>28.4</td>
<td>25.1</td>
<td>21.3</td>
</tr>
<tr>
<td>group interaction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>experiential</td>
<td>26.4</td>
<td>20.7</td>
<td>13.0</td>
</tr>
<tr>
<td>reflective</td>
<td>28.0</td>
<td>28.1</td>
<td>25.6</td>
</tr>
<tr>
<td>controls</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>practice-only</td>
<td>27.4</td>
<td>21.6</td>
<td>19.9</td>
</tr>
<tr>
<td>predict-only</td>
<td>31.9</td>
<td>27.9</td>
<td>28.0</td>
</tr>
</tbody>
</table>

*Note.* Means are expressed in hundreds of tons. For example, a mean of 28.0 means an average deviation of 2800 tons off target. The rightmost column shows correlations between final policy quality and final performance.

* indicates \( p < .05 \).
It was predicted that a combination of assisted reflective practice and exposure to alternative models was necessary for changing peoples' mental models of the task (as reflected in their written policies). That is, an interaction between exposure to other’s policies and practice mode was expected. Although the interaction between practice mode and exposure to other’s policies was significant, $F(2,109) = 5.76$, MSE = 106.1, $p = .004$, its direction was not as expected. The overall effect of assisted reflective practice was negative, $F(1,109) = 10.06$, MSE = 106.1, $p = .002$. Test of simple main effects revealed that the effect of practice mode was only present for the group interaction condition where the exposure to alternative models was present and reinforced through group discussion, $F(1,109) = 21.07$, MSE = 106.1, $p < .01$.

Contrary to these results, Mathews et al. (1996) reported that reflective practice led to large improvements in policy quality for participants in their groups condition. A way to reconcile this difference in results is to examine the information available to the participants during reflective practice.

In Mathews et al. (1996), feedback to participants about policy quality was generated by simulating the policies with a computer program. To prepare them for simulation, the policies were first translated into sets of condition-action rules (e.g., if production is high, use...
fewer workers). Policy statements that were ambiguous were not included in the simulation. The simulation program was a variant of a classifier system (Holland et al., 1986; Druhan & Mathews, 1989; Roussel & Mathews, 1990), which is a production system in which rules compete against each other based on strength (past success rate for each rule), specificity (number of conditions specified), and support (agreement in action with other rules) to control the system's response. In this way, the simulation not only evaluates the policy by using it to control sugar production, but it also adjusts the strength of the rules to optimize success with a given set of rules (Mathews et al., 1996).

The feedback received by the participants in Mathews et al. (1996) was a score indicating the success of the optimized policy controlling sugar production. Additionally, any rules whose final strength at the end of the simulation was below a threshold value (indicating they never succeeded in the rule competition process in the simulation) were indicated as such by a note written directly on the policy. Thus participants received information about which rules were ineffective. This allowed learners to focus only on the most valid rules. By contrast, in the current experiment, participants were given only a global policy score indicating the quality of the policy as a whole.
In addition to informing them about the relative effectiveness of each of their rules, the reflective practice procedure of Mathews et al. (1996) directed the participants' attention toward the best rule on each trial. On each trial the participants were required to predict how many workers their policy would use and which particular rule would be employed (by the simulation program) to set the size of the workforce. Following their responses, they received feedback on what workforce size their policy would have set in this situation and which rule was used.

As a result of these differences between the reflective practice procedures of Mathews et al. (1996) and the current experiments, participants in the experiment of Mathews et al. probably gained more experience with more valid rules and less experience with invalid rules. Further, the more extensive feedback concerning the relative effectiveness of individual rules given by Mathews et al. (1996) may have helped their participants to improve their reflective knowledge by eliminating ineffective rules.

A qualitative look at policies in the group interaction conditions in the current experiment showed that reflective knowledge was drastically different across reflective and experiential conditions. The policies of participants performing experiential practice consisted largely of statements of specific situations (e.g., when production is
3000 tons, use 500 workers) whereas policies of participants performing reflective practice rarely contained specific situations. The mean number of specific situations mentioned in the experiential and reflective conditions were 7.21 (SD = 5.26) and .95 (SD = 2.67), respectively. Typical policy statements in the reflective practice condition were elaborations of the (invalid) belief that more workers produce more sugar (e.g., if production is below target, then add 50 to 100 workers). Almost as frequent were references to patterns of sugar output across several consecutive trials and the corresponding response. For example, "if production changed from high to low, then production level will probably increase this trial, therefore use less workers to compensate for the expected increase."

These differences in participants' conceptions of the sugar production task show that assisted reflective practice did have a powerful influence on reflective knowledge, even if that influence was negative. Although it was contrary to the predicted effect, it is consistent with findings of implicit learning studies in which rule search fails (e.g., Reber 1976; Reber et al., 1980). These results, along with the interpretation of the differences between the reflective practice task of Mathews et al. and assisted reflective practice also suggest that presenting learners with a few valid and specific situations might be...
beneficial. This possibility is taken up in subsequent experiments.

The Effect of Reflective Practice on Task Performance

It is assumed that experiential knowledge is best measured by the ability to control sugar production. Thus, task performance during each test session is taken to be an index of experiential knowledge. Table 1 displays the mean performance in each condition across the three test sessions.

Performance means were analyzed within a repeated measures ANOVA. The three factors included in the ANOVA were session, practice mode, and exposure to other's policies. Session was the repeated factor. The two control conditions, practice-only and predict-only were excluded from this analysis and are addressed later.

Performance improved across sessions, \( F(2,218) = 37.44, \text{MSE} = 41.6, \ p < .0001 \), indicating that participants were gaining experiential knowledge. Participants in experiential conditions consistently outperformed their reflective counterparts, \( F(1,109) = 10.90, \text{MSE} = 137.8, \ p < .001 \), contrary to the expected benefit of reflective practice. The interaction between practice mode and session was also statistically significant, \( F(2,218) = 8.98, \text{MSE} = 41.6, \ p < .0001 \). The pattern of means suggested that the negative effect of reflective practice increased across trials. To interpret this interaction, tests of simple
main effects were performed to compare reflective against experiential at each level of session. The error term for these tests was formed by pooling the between-subjects and within-subjects error terms (see, Kirk, 1982, for an explanation of the procedure) and the significance level (alpha) was adjusted using Bonferroni's procedure. The practice mode effect was not present at session 1, $F(1,327) = .005$, $MSE = 73.65$. This was expected because participants in the reflective and experiential conditions were treated identically at the first session. The effect of practice mode was statistically significant at sessions 2 and 3, $F(1,327) = 11.36$, $MSE = 73.65$, $p < .01$, and $F(1,435) = 19.74$, $MSE = 73.65$, $p < .01$, respectively.

The next analysis of performance focused on the final session where one would expect any treatment effects to be the greatest. Data from sessions 1 and 2 were excluded from this analysis. These results also can be directly compared to the findings for reflective knowledge discussed above.

Participants in the experiential conditions outperformed participants in the reflective conditions, $F(1,109) = 16.05$, $MSE = 87.9$, $p = .0001$. There was no main effect of exposure to other's policies, $F(2,109) = .19$, $MSE = 87.9$, $p > .05$, but the interaction between practice mode and exposure to other's policies was significant, $F(2,109) = 3.20$, $MSE = 87.9$, $p = .045$. Tests of simple
main effects showed that the practice mode effect was marginal for the no exposure condition, $F(1,109) = 5.58$, $MSE = 87.9$, $p = .03$, nonsignificant for nominal groups, and significant for the group interaction condition $F(1,109) = 17.3$, $MSE = 87.9$, $p < .01$. Inspection of the means shows the effect of practice mode largest in the group interaction condition ($M = 13.04$ vs $M = 25.57$ for experiential and reflective conditions, respectively).

Other than the marginal effect of reflective practice for the no exposure condition, the results for experiential knowledge were the same as those for reflective knowledge. Thus, the same positive effects of group discussion and experiential practice were found for reflective knowledge and experiential knowledge.

The final analysis of experiential knowledge investigated differences between the two control conditions and the no exposure/experiential condition. Recall that participants in the practice-only control condition were treated the same as those in the no exposure/experiential condition, except that the practice-only controls did not write a policy at the end of each session. Thus, a comparison between these two conditions will reveal any effect of post-task reflection in the form of writing a policy. Participants in the predict-only control condition also did not write a policy. However, the prediction task can be considered as an unassisted reflective practice.
task. Thus, the comparison between the two control conditions will allow an assessment of the effect of unaided reflection.

The three conditions differed in final session performance, $F(2, 57) = 6.04$, MSE $= 105.0$, $p = .004$. Tukey/Kramer pairwise comparisons revealed that policy writing had no effect on performance; the difference between practice-only and no exposure/experiential was not significant, $Q(57) = 1.35$, $p > .05$. However, the requirement to predict sugar production on each trial significantly hurt performance relative to the practice-only controls, $Q(57) = 3.42$, $p < .05$. Thus, unaided reflection actually hurt performance.

It is possible that the consistent negative effect of reflection across all conditions arose out of differences in the number of practice trials experienced by the experiential and reflective learners. Because reflective practice requires slow and deliberate responding, learners in reflective conditions might simply have experienced fewer practice trials. An examination of data from practice sessions showed that experiential and reflective learners completed an average of $1092.8$ ($SD = 413.0$) and $440.9$ ($SD = 237.9$) trials, respectively. A correlation analysis was performed to ascertain the contribution of accumulated practice to performance on the final test. The results showed that the number of practice trials was
significantly correlated with test performance, \( r = -.283, \ p < .05 \). However, the number of practice trials is confounded with practice mode. Therefore, a semi-partial correlation coefficient was computed in which the correlation between test performance and practice mode was removed from the correlation between test performance and number of practice trials. The resulting semi-partial correlation coefficient was essentially 0, \( r = -.022, \ p > .05 \). Therefore, the poor performance of reflective learners does not appear to be due to completion of fewer practice trials.

The Effect of Providing Alternative Models

As already noted, there were no main effects of exposure on either experiential or reflective knowledge. The provision of alternative models did interact with mode of practice. However, tests of simple main effects (reported above) showed that the interaction was explained by a large practice mode effect in the group interaction condition. The provision of alternative models alone (i.e., in the absence of group interaction) had no effect on performance.

The Effect of Group Discussion

As just discussed, the interaction between exposure to alternative models and practice mode was accounted for by the practice mode effect in the group interaction condition. Thus, group interaction did have some effect.
beyond that of exposure to alternative models. Recall also that the final policies of experiential and reflective learners in the group interaction condition were distinguished by a difference in the number of specific situations mentioned. Relative to the policies of reflective learners, the final policies of experiential learners tended to contain more statements referring to specific situations. This finding suggests that group discussion (post-task reflection) played some role in the development of learners' mental models. To examine this role, a simple qualitative analysis of policies was carried out.

To get a clear picture of the effect of group discussion, a comparison was made between the policies of the group interaction condition and the nominal groups condition. These conditions differed only by the presence or absence of group interaction. The policies chosen for examination were the group policies written at the beginning of the second practice session. Up until the time these policies were written, participants in the two conditions had been treated identically except for the group policy formation task (See Figure 1). Recall that each participant in the nominal groups condition formed his or her own group policy after being given copies of the policies of the other group members. By contrast, the participants in the group interaction condition met as a
group to construct a single group policy. Because at this stage of the experiment experiential and reflective learners had been treated identically, the data from these two conditions were pooled for this analysis.

The number of specific situations mentioned for the nominal groups and group interaction conditions were tallied as before. To determine if the groups differed in their tendency to generate specific statements, each policy was categorized as follows: A general policy, if no specific situations were mentioned, or a specific policy, if one or more specific situations were mentioned. One-third of the policies in the nominal groups condition were categorized as specific policies compared to two-thirds of the policies of the group interaction condition. This difference was marginally significant, $\chi^2 = 3.334, p = .072$.

Although not strong, this evidence suggests that post-task reflection might play a role in the interaction between practice mode and exposure to alternative models. Possibly, group discussions lead to more specific and better quality policies but reflective practice can subvert the process. Conversely, experiential practice may interact positively with the more specific policies to lead to good performance.
Integration of Reflective and Experiential Knowledge

As a way of assessing integration between reflective and experiential knowledge and as check on the validity of the procedure for measuring policy quality, correlations between policy quality and performance were computed. Large correlations cannot prove that there was integration (agreement) between experiential and reflective knowledge nor can it prove that the measurement of policy quality was valid. However, very small correlations would indicate either lack of integration or lack of validity. The correlations are displayed in the rightmost column of Table 1.

Correlations were large and statistically significant for all except the nominal groups/reflective condition. A scatter-plot for the nominal groups/reflective condition was examined to check for any spurious data that might explain the lack of correlation. That check revealed that the small correlation was not accounted for by any spurious data points.

These correlation results suggest that learners were able to integrate their experiential and reflective knowledge of the task. However, this may sometimes hurt rather than help performance (e.g., in the group interaction/reflective practice condition). The one nonsignificant correlation in the nominal groups/reflective practice condition indicates that learners in this
condition were unable to make sense of the wealth of information they were provided with.

The results of experiment 1 suggested that the specificity of learners' mental models was related to their ability to control sugar production; the policies of participants in the best performing condition (experiential/group interaction) tended contain many statements of specific situations. Further, reflective practice led learners to generate very general rules of low validity. Experiment 2 examined this relationship by providing some learners with hints consisting of specific or concrete situations illustrating how to control sugar production. The question addressed was whether providing these hints would prevent reflective practice from causing learners to generate overly general policies.
EXPERIMENT 2

Introduction

One possible explanation for the negative effect of assisted reflective practice in Experiment 1 is a tendency to seek or create only general rules, which are not very useful in performing the task. I refer to this tendency as reflective abstraction error. Perhaps participants in the high performing group interaction/experiential practice condition avoided the error by not reflecting while performing the task. Consequently, the group discussion was focused more on concrete experiences or cases in which they successfully controlled sugar production rather than abstract rules about the relations between workers and production.

Experiment 2 attempted to inhibit participants' tendency to create overly general rules (i.e., reflective abstraction error) by giving them a hint describing concrete or specific situations. The hint comprised valid work force levels for four specific situations. It was presented to some participants at the beginning of the second and third practice sessions. It was predicted that the hint would induce learners to switch to a strategy of encoding the task as specific condition-action rules in a manner similar to participants in the group interaction/experiential condition of Experiment 1. The consequence of encoding the task as specific situations
would be to reduce the chance for reflective abstraction error and thus lead to a corresponding improvement in performance.

The hints can be construed also as a variation of the procedure in Experiment 1 for providing learners with an alternative model of the task. The difference is that the hints are provided by the experimenter and are valid. By contrast, the alternative models of Experiment 1 were provided by other participants and were likely to be invalid. Thus, Experiment 2 provides a more powerful test of the effect of exposing participants alternative models of the task.

Method

Participants and Design

Fifty-two undergraduate students enrolled in introductory psychology courses at Louisiana State University were recruited to voluntarily participate in return for extra-credit.

Three experimental conditions were examined: hint/experiential practice, hint/assisted reflective practice, and a no hint/experiential practice control condition. As in the previous experiment, session was a within-subjects factor.

The primary dependent variable was performance as indicated by the average unsigned deviation from target
production. An additional dependent measure, policy quality, was assessed as it was in Experiment 1.

**Procedure**

Participants were tested in groups of three to six individuals. Each group was randomly assigned to one of the three conditions. Regardless of condition, all participants completed three sessions, one per day. For all participants, the three sessions were completed within seven days. The procedure was identical to that followed in Experiment 1 for the no exposure conditions except for the provision of the hints described below. That is, all participants were exposed only to their own policies or to an experimenter-provided hint. Participants did not discuss their polices with others in any conditions.

Participants in the hint conditions received a hint in the form of a typed text at the beginning of the second and third sessions. They were told that the hint would help them learn to control sugar production. The hint was taken away at the beginning of the test phase for each session. The hint was as follows:

The number of workers should always follow the level of production. That is, when production is high, you need a lot of workers and when production is low, you need few workers. Similarly, when production is near the middle, you should use a moderate level of workers, not high and not low. Below are four specific examples.

If production is 1000 tons, then use 350 workers.
If production is 4000 tons, then use 500 workers.
If production is 7000 tons, then use 650 workers.
If production is 10000 tons, then use 800 workers.
Results and Discussion

The Effect of the Hint on Strategy

The number of specific situations mentioned in each participant's policy was tallied. A median test for independent groups (Siegel & Castellan, 1988) revealed that there were differences across conditions in the number of specific situations mentioned, $\chi^2(2) = 17.79, p < .05$. The mean number of specific situations mentioned was 1.96, 5.82, and 3.95, for the control, hint/experiential, and hint/reflective conditions, respectively. The two hint conditions did not differ from each other, $\chi^2(1) = .017, p > .05$, but both differed from the control condition, $\chi^2(1) = 12.78, p < .05$.

The Effect of the Hint on Reflective Knowledge

Although provision of the hint did cause participants to mention more specific situations in their policies, this does not guarantee that the resulting policies were more valid. To answer that question, the final policies for all participants were simulated as in Experiment 1. The mean performances for the simulated final policies are displayed in Table 2.

The performance of the simulated policies differed across the three conditions, $F(2,49) = 10.37$, $\text{MSE} = 145.8$, $p = .0002$. However, only the hint/experiential condition was better than controls, $Q(49) = 6.34, p < .05$. The hint/reflective condition did not differ from the control condition.
Table 2. Mean absolute deviation from target by condition and session in Experiment 2.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Session</th>
<th></th>
<th></th>
<th>Final Policy</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>no hint/experiential</td>
<td>30.0</td>
<td>26.2</td>
<td>21.5</td>
<td>29.55</td>
<td>.24</td>
</tr>
<tr>
<td>hint/experiential</td>
<td>28.6</td>
<td>9.4</td>
<td>6.9</td>
<td>10.68</td>
<td>.60*</td>
</tr>
<tr>
<td>hint/reflective</td>
<td>29.9</td>
<td>14.7</td>
<td>10.7</td>
<td>24.54</td>
<td>.31</td>
</tr>
</tbody>
</table>

Note. Means are expressed in hundreds of tons. For example, a mean of 28.0 means an average deviation of 2800 tons off target. The rightmost column shows correlations between final policy quality and final performance.

* indicates p < .05.
condition, $Q(49) = 1.81$, $p > .05$. Thus, assisted reflective practice had the same destructive effect on reflective knowledge as it did in Experiment 1.

The Effect of the Hint on Task Performance

The three conditions differed overall, $F(2,49) = 8.91$, $MSE = 182.2$, $p = .0005$ and performance improved across sessions, $F(2,98) = 93.34$, $MSE = 40.6$, $p < .0001$. The interaction between condition and session was also statistically significant, $F(4,218) = 8.98$, $MSE = 40.6$, $p < .001$. Tests of simple main effects showed that condition was not significant at the first session. This was expected because participants in all conditions were treated the same. However, performance differed across treatments for the second, $F(2,147) = 14.68$, $MSE = 87.8$, $p < .01$, and third session, $F(2,147) = 11.50$, $MSE = 87.8$, $p < .01$.

Participants in both hint conditions performed better than controls during the final session. The difference between the hint/experiential and hint/reflective conditions was not significant, $Q(49) = 1.612$, $p < .05$. Recall that in Experiment 1 reflective practice reduced both the final level of task performance and the final policy quality, relative to experiential practice. Providing the hint in Experiment 2 overcame the negative impact of reflective practice on task performance but not on final policy quality. This result suggests an increased
dissociation between experiential and reflective knowledge under conditions of reflective practice.

Further evidence of this dissociation is the lack of a significant correlation between final policy quality (reflective knowledge) and final test performance (experiential knowledge). Reflective and experiential knowledge were correlated in the experiential/hint condition, \( r = .60, p < .05 \), but not in the reflective/hint condition, \( r = .31, p > .05 \). The lack of a significant correlation in the no hint control condition is puzzling in that this condition was identical to the no exposure/experiential condition of experiment 1 where a significant correlation was found. There is no explanation of this inconsistency between experiment 1 and experiment 2.

The next experiment explored this dissociation. Specifically, the reflective practice procedure was modified in an attempt to overcome the dissociation between experiential and reflective knowledge. Additionally, the next experiment incorporated a transfer task in order to investigate the flexibility of knowledge in the sugar production task.
EXPERIMENT 3

Introduction

In Experiment 2, the hint improved performance but did not improve policy quality when learners practiced reflectively, suggesting an increased dissociation between reflective and experiential knowledge. In response to that finding, the reflective practice procedure of Experiments 1 and 2 was replaced by a simpler method for fostering reflective practice. In Experiment 3, participants in the reflective practice conditions were provided with paper and pencil during performance of the task and encouraged to record their behavior and its results while performing the task. They were advised to use their notes to help them with their decisions about workforce size. The use of cognitive artifacts like pencil and paper is characteristic of reflective thought (Norman, 1993).

It was expected that the use of pencil and paper would cause participants to "stop and think" as they worked (Gagne & Smith, 1962). Further, it would serve as external memory and thus lessen the demand for cognitive resources as they worked. Participants in the experiential conditions were not allowed to use pencil and paper thereby removing support for reflective thought. It was assumed that removing this support would lead to a default to experiential processing.
Experiment 3 also addressed the issue of flexibility of knowledge. Some researchers have suggested that implicitly (experientially) acquired knowledge is inflexible. For example, Dienes and Fahey (1995) found that performance was at chance levels for situations which had not been experienced previously. That is, transfer of knowledge was highly specific (Berry & Dienes, 1993).

In Experiment 3, a transfer task was used to investigate flexibility of knowledge. During the final test phase, participants were asked to achieve a new (novel) target production on half of the trials. If experiential knowledge is as specific as claimed then performance on new targets should be near chance.

Because the hint had such a powerful effect on performance in Experiment 2, it was included here to investigate its effect on transfer. Two levels of hint were examined, a general hint that mentioned only relative values for workers and production, and a specific hint that gave four specific situations in addition to the general hint for controlling sugar production. It was expected that the general hint would show the most transfer to the new target level. The specific hint applied only to the old target level and thus it was not clear what effect it would have on performance during transfer.
Method

Participants and Design

One hundred twenty-one undergraduate students enrolled in introductory psychology courses at Louisiana State University were recruited to voluntarily participate in return for extra-credit. As before, all participants were tested in groups of three to six individuals.

Three independent variables were arranged in a factorial design: practice mode (reflective vs experiential), hint (none vs general vs specific), and target level (old vs new). The no hint/experiential condition was considered to be a control condition. Target level was a within-subjects factor. The repeated factor of session was also included as in the previous two experiments.

Procedure

Each group of three to six individuals was randomly assigned to one of the six conditions. The general procedure was identical to that followed in Experiment 2 in that each session consisted of a practice phase and a test phase. However, the original, or old, target level was exchanged for a new target level for half of the trials during the test phase of the third session.

Two target levels were used: 4000 and 8000. Each individual was randomly assigned to one of the target levels for the first two sessions. This was considered the
"old" target level. Midway through the test phase of the final session, the target level was switched to the other target for the remaining trials. Participants were warned about the change and were told to do their best at achieving the new target level. Thus, each participant performed twenty trials with each target level.

During the test phase of each session, participants in the reflective conditions could use any notes they had made during the practice phase. Thus, any superiority of the reflective conditions could be due to having this aid during the test and not due to reflection during practice. To control for this advantage, an extra test phase was included at the end of the experiment. At the end of the third session, and after all participants had written their final policy, all notes and policies were taken from the participants and they completed twenty more test trials. During this test, the same two target levels were used as before. The old target level was used for the first ten trials and the previously new target level for the final ten trials.

**Results and Discussion**

**The Effect of Reflection (Paper and Pencil)**

Mean test performance across sessions (collapsed over old and new target levels) is displayed in Table 3. There were no differences between the experiential and reflective conditions, $F(1,115) = .25$, $MSE = 139.9$, $p = .62$, nor was
Table 3. Mean absolute deviation from target by condition and session, collapsed over target level for Experiment 3.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Session</th>
<th></th>
<th></th>
<th>Final Policy</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td>£</td>
</tr>
<tr>
<td>no hint</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>experiential</td>
<td>23.7</td>
<td>20.1</td>
<td>21.6</td>
<td>35.5</td>
<td>.14</td>
</tr>
<tr>
<td>reflective</td>
<td>25.2</td>
<td>22.3</td>
<td>20.6</td>
<td>33.0</td>
<td>.05</td>
</tr>
<tr>
<td>general hint</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>experiential</td>
<td>25.6</td>
<td>14.0</td>
<td>16.0</td>
<td>31.3</td>
<td>.06</td>
</tr>
<tr>
<td>reflective</td>
<td>24.6</td>
<td>12.6</td>
<td>15.8</td>
<td>36.0</td>
<td>-.04</td>
</tr>
<tr>
<td>specific hint</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>experiential</td>
<td>20.3</td>
<td>11.7</td>
<td>15.4</td>
<td>31.8</td>
<td>.18</td>
</tr>
<tr>
<td>reflective</td>
<td>24.9</td>
<td>12.3</td>
<td>15.6</td>
<td>32.7</td>
<td>-.01</td>
</tr>
</tbody>
</table>

Note. Means are expressed in hundreds of tons. For example, a mean of 28.0 means an average deviation of 2800 tons off target.
the interaction between practice mode and hint significant, $F(2,115) = .41$, $MSE = 139.9$, $p = .67$. It is quite surprising that being allowed to record one's interactions with the task had no effect on performance. One would expect that, at least on occasion during the test, a participant would encounter a situation similar to one experienced during practice and could therefore benefit from a written record of that experience. But, obviously, this did not happen. Indeed, the same analysis performed on the simulated final policy performance (Table 3) revealed no main effects or interactions for practice mode or hint (all $Fs < 1$). It seems the ability to keep a written record did not influence policies either.

Although reflection had no effect on either policy quality or performance, it might have influenced the dissociation between experiential and reflective knowledge found in the previous experiment. The correlations between experiential and reflective knowledge (Table 3) are all small and not significantly different from zero. This would seem to suggest that there was a strong dissociation present across all conditions. However, the design of the transfer task allowed participants' policies to be ambiguous with respect to the target level to which they applied. Recall that the final polices were written after the final test. In the final test, target levels were changed in order to investigate transfer. Subsequently,
when participants wrote their final policies, most did not state whether their policies applied to old or new targets, or both. Neither did they ask the experimenter whether they should specify target levels in their policies. Thus, the low correlations are not necessarily indicative of a dissociation.

The Effect of Hints on Task Performance

The effect of hint type was significant, $F(2, 115) = 7.13$, $\text{MSE} = 139.9$, $p = .0012$, and interacted with the session variable, $F(4, 230) = 5.73$, $\text{MSE} = 39.6$, $p = .0002$. The effect of hint was not present at session 1 ($F<1$), but was significant for session 2, $F(2, 345) = 13.67$, $\text{MSE} = 73.0$, $p < .01$, and for session 3, $F(2, 345) = 5.27$, $\text{MSE} = 73.0$, $p < .01$. Both hint conditions were superior to controls (no hint) at session 3 but did not differ from each other.

An effect of hint type on experiential knowledge (performance) without a corresponding effect on policy quality is consistent with the differential impact of hint in the hint/reflective condition of experiment 2. This demonstrates that learners draw upon distinct knowledge for performing and for communicating (i.e., writing policies).

Flexibility of Knowledge

Because practice mode had no effect on performance, data from experiential and reflective conditions were combined for the remaining analyses.
Flexibility was assessed by measuring performance separately for old and new target levels. Mean performances for old and new targets are presented in Table 4 under the heading "first transfer test." Additionally, the difference between performance on old and new targets is also displayed. Data from the first transfer test are considered first.

Performance was better for old targets than for new, $F(1,118) = 91.88$, MSE = 605.3, $p = .0001$. Looked at this way, there was negative transfer to new targets. However, performance on old targets is not the proper baseline for measuring transfer. One way to measure transfer is to compare performance on new targets relative to chance performance.

To ascertain whether performance on new targets was better than chance, 99 percent confidence intervals were constructed around performance means for new targets in each condition. These confidence intervals were inspected to see if they contained 42, which is chance performance. The upper bounds for all three confidence intervals were below 30 (hundred tons) which is well below 42, the deviation score achieved by random selection of worker levels. This indicates that performance on new targets was better than chance.

Hint level also interacted with target level, $F(2,118) = 4.58$, MSE = 605.3, $p = .012$. To interpret the
Table 4. Mean absolute deviation from target by type of hint and target before and after retention interval.

<table>
<thead>
<tr>
<th>hint</th>
<th>first transfer test</th>
<th>second transfer test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>old</td>
<td>new</td>
</tr>
<tr>
<td>none</td>
<td>17.5</td>
<td>24.7</td>
</tr>
<tr>
<td>general</td>
<td>12.0</td>
<td>19.7</td>
</tr>
<tr>
<td>specific</td>
<td>8.6</td>
<td>22.4</td>
</tr>
</tbody>
</table>

Note. Means are expressed in hundreds of tons. For example, a mean of 28.0 means an average deviation of 2800 tons off target. The values in the columns labeled "old-new" indicate the difference in performance on old and new targets.
interaction, Tukey/Kramer pairwise comparisons were made separately for old and new targets. The performance means on old targets were 17.5, 12.0, and 8.6, for the no hint, general hint, and specific hint condition, respectively. Both hint conditions were better than the no hint condition. Although the mean for the specific hint condition was lower than the mean for the general hint condition (8.6 vs 12.0), that difference was not significant. Thus, hints were effective in improving performance, at least on old targets.

The performance means for new targets were 24.7, 19.7, and 22.4, for the no hint, general hint, and specific hint condition, respectively. No differences between the three conditions were significant. Thus, neither hint supported transfer to new targets during the first transfer test.

I now turn to the second transfer test. As in the first transfer test, the effect of target was significant, with performance on old targets being superior to performance on new targets, $F(1,118) = 55.34$, $MSE = 709.2$, $p = .0001$. However, the interaction between target level and hint type was not present at the second transfer test, $F < 1$.

Tukey/Kramer pairwise comparisons computed for each target level revealed that the specific hint condition was superior to the no hint condition for both old and new targets. Thus, participants provided reflective knowledge
in the form of a specific hint, were able to apply the hint to the new targets. The general hint condition did not differ from either the no hint condition or the specific hint condition.

Although the new target level is not truly "new" during the second transfer test (i.e., the new target level was experienced during the first transfer test), the superiority of the specific hint condition is surprising. After all, the specific hints only applied to the old target level and if followed for the new target would lead to poor performance. Apparently, the minimal experience learners got with the new target during the first transfer test enabled them to apply the hints to the new targets during the second transfer test.
GENERAL DISCUSSION

This investigation examined the effects of reflection, exposing learners to others' ideas about the task, and the opportunity to discuss those ideas with other learners. An additional concern of this investigation was the effect that reflection would have on the integration of reflective and experiential knowledge. In three experiments, the effectiveness of various kinds of reflective thinking, involving both within-task and post-task reflection, was investigated. Also, two ways of providing learners with alternative ideas about task behavior, exposure to other's policies and providing hints for task solution, were examined. Following a summary of the basic findings regarding knowledge integration, possible explanations and the theoretical implications will be discussed.

Summary of Findings

Within-task Reflection

Several methods for eliciting within-task reflection were investigated. The simplest method was to give learners pencil and paper along with instructions to use them as they saw fit to help them learn the task (Experiment 3). It was expected that pencil and paper would serve as external memory and lessen the demand for cognitive resources. The use of pencil and paper was also expected to cause participants to "stop and think", or
perform more reflectively, as they worked (Gagne & Smith, 1962).

Contrary to these expectations, no effect on learning was found. Performance and policy quality were the same in Experiment 3 whether participants had paper and pencil or not. This result confirms what others have argued is a tendency for people to default to an experiential mode when reflective strategies meet with difficulty (e.g., Berry & Dienes, 1993; Reber, 1993).

Another method for eliciting within-task reflection during task performance was to require participants to predict the outcome of each selection of workforce size (prediction condition, Experiment 1). This method apparently was successful in eliciting reflection because the performance of participants in the prediction condition differed from that of participants in a non-prediction control condition. However, this kind of reflection led to worse rather than better performance. Thus, forcing learners to reflect as they learn can actually interfere with learning.

The final method for eliciting within-task reflection was assisted reflective practice which involved a computer program designed to assist learners in thinking about their policies for controlling sugar production and to help them evaluate their policies by using them to perform the task (Experiments 1 and 2).
Assisted reflective practice was found to be quite damaging to learning and performance. The policies created by learners engaging in assisted reflective practice were less valid than those of learners not required to reflect as they practiced. Also, at the end of training, task performance was worse following assisted reflective practice compared to experiential (non-reflective) practice.

In summary, eliciting within-task reflection failed to improve learning in every condition investigated in these experiments. Contrary to my predictions, within-task reflection consistently led to worse performance relative to participants allowed to perform the task experientially. Further, there was evidence that within-task reflection led to less valid reflective knowledge. The simulated policies of participants in reflective conditions tended to perform worse than those of participants in experiential conditions.

That supported reflection could be so detrimental to learning was unexpected considering the number of studies finding beneficial effects of a variety of reflective processes (e.g., Ahlum-Heath & DiVesta, 1986; Berardi-Colleta et al., 1995; Biemiller & Meichenbaum, 1992; Chi et al., 1987, 1994; Gagne & Smith, 1962; Klahr & Dunbar, 1988; Mathews et al., 1996; McGeorge & Burton, 1989; Stanley et al., 1989; Trudel & Payne, 1995; Wilder & Harvey, 1971).
The size and robustness of this negative impact of reflective practice in these experiments justifies the creation of a name for this phenomenon: reflective abstraction error. The theoretical implications of reflection abstraction error are taken up later in this discussion.

**Post-task Reflection**

Two kinds of post-task reflection were investigated. In one kind, participants wrote down their strategies for controlling sugar production at the end of each practice session. In the other kind of post-task reflection, participants met periodically in small groups to discuss their written strategies and create a group policy for controlling sugar production.

The requirement to write a strategy had no effect on performance. However, participation in group discussions had a significant effect on learning that varied according to whether within-task reflection was present. This interaction effect can be seen by comparing the nominal groups and group interaction conditions of Experiment 1. In the nominal groups condition, participants received copies of the strategies of other group members. They were instructed to combine those policies, along with their own policy, into a single policy. Then they were instructed to evaluate this policy during subsequent practice sessions. Nominal groups participants were not
allowed to discuss the policies with the other group members contributing the policies. In the group interaction condition, participants were allowed to see and evaluate each other's policies and to discuss their policies with the goal of creating the best single group policy for controlling sugar production. Thus, they had the greatest opportunity for reflecting about their own and other's ideas.

The results showed that when participants were allowed to discuss their policies with other members, those who had assisted reflective practice performed much worse than those who had experiential practice. By contrast, in the nominal groups conditions (no group discussion), performance of those having assisted reflective practice did not differ from the performance of those having experiential practice. In addition, performance of the group interaction participants who had experiential practice exceeded that of participants in the nominal groups/experiential practice condition. This was the only condition in which reflection (group discussion) actually improved task performance.

These results suggest that the effect of post-task reflection (discussion) depends on the content of the discussion. The policies of participants in the experiential practice conditions contained significantly more examples of specific situations than the policies of
participants in the assisted reflective practice conditions. Thus, the group discussions of experiential learners were more focused on specific situations encountered in the context of the task.

Exposure to Other's Policies

Another variable investigated for its effect on experiential and reflective knowledge was exposure to alternative task conceptions. It was hypothesized that conceptual change would be most likely to occur when learners have access to a variety of task conceptions.

Although access to alternative task conceptions, through exposure to other's policies, was predicted to lead to superior performance when combined with assisted reflective practice, it actually had no effect. In Experiment 1, there was little difference in performance between participants who had access only to their own policies (no exposure condition) and those who were allowed to evaluate other's policies (nominal groups condition). This finding suggests either that learners ignore ideas not generated by themselves or that they are unable to integrate those ideas into their own conception of the task. The lack of correlation ($r=-.02$) between performance and policy quality in the nominal groups condition suggests that participants were unable to integrate the policies of others with their own experiential knowledge. That is, their experiential knowledge, which controlled task
performance, became dissociated from their reflective knowledge which was influenced by exposure to others' policies.

However, when groups of participants were allowed to discuss each others policies the dissociation between experiential and reflective knowledge was eliminated. That is, in the group discussion conditions, experiential and reflective knowledge were significantly correlated and were affected similarly by manipulations of practice mode (experiential vs reflective practice). Further, combining group discussion with exposure to other's policies actually facilitated task performance for participants practicing experientially.

**Giving Learners Valid Hints**

Only when participants were given valid hints about controlling sugar production (Experiment 2) did assisted reflective practice not significantly interfere with learning. Learners given hints about how to control sugar production performed near perfect by the final session and there was no difference in performance between experiential and assisted reflective practice. Apparently, when within-task reflection is directed toward valid cases (hints), the damaging effect of reflective practice is eliminated, or at least greatly reduced. However, even though performance was not negatively affected, learners performing assisted reflective practice still tended to develop less valid
mental models as indicated by their poorer quality policies.

Giving learners hints also provided an opportunity to investigate the relative transfer of reflective knowledge provided in the hints, and experiential knowledge acquired from task experience. Experiential knowledge is commonly viewed as being inflexible and tied to specific perceptual cues (Berry & Dienes, 1993; Schacter, 1987). In particular, positive transfer to new situations in the sugar production task is rare (Dienes & Fahey, 1995). By contrast, reflective knowledge is generally viewed as flexible and applicable to novel situations (Baars, 1988). The results of Experiment 3 suggest that this view may be too strong.

In Experiment 3, at the beginning of the second practice session, some participants were given hints specifying the correct response for four different situations. Those hints were valid for the target level used during practice sessions. However, at the final session the target level was changed for the transfer test, thereby rendering the specific content of the hints (e.g., if production is 3000 then use 400 workers) invalid. Clearly, participants could not literally apply the hints to perform the transfer test. They had to abstract or adapt this knowledge in order to apply it to the new target level. Thus, this transfer test taps a very abstract level
of knowledge, or at the least, measures the use of abstract analogy during test.

In the first transfer test, participants given the hints performed no better on new targets than participants in the no-hint control condition. This suggests that reflective knowledge (i.e., knowledge acquired from hints) did not facilitate transfer. However, performance on new targets in both the hint and no-hint conditions was still better than chance performance. This suggests that the experiential knowledge learners gained from task experience did transfer to the new target level.

After a brief retention interval, during which all participants' written policies and notes were taken away, a second transfer test was given using the same new target levels as for the first transfer test. Of course, the new target levels are no longer truly new as they were seen during the earlier transfer test. However, participants had only minimal experience (20 trials) with the new targets.

Interestingly, during the second transfer test, participants that had received the specific hints performed better on new targets than the no-hint condition. Remember, these hints were not literally valid for new targets. Apparently, the minimal experience that participants received with the new targets during the first transfer test was sufficient for them to modify their
reflective knowledge and integrate it with their experiential knowledge, increasing their performance on the second transfer test. By contrast, participants who received only general hints (i.e., hints not referring to specific target levels) performed no better than no-hint controls on the second transfer test.

Knowledge Integration

Several investigations using the sugar production task and similar tasks have reported dissociations between experiential and reflective knowledge (e.g., Berry & Broadbent, 1984, 1988; Stanley et al., 1989). One of the questions addressed in this investigation was whether engaging reflective thinking in the context of performing the task would lead to greater integration (less dissociation) of experiential and reflective knowledge.

The results of these experiments suggest that integration is most likely to occur when two conditions are met: 1) Reflection occurs following task performance in the context of discussions with other learners about performing the task. 2) Reflection is focused on specific cases or direct experiences rather than abstracted theories or beliefs about the task. When these two conditions are not met, either performance will suffer, or dissociation will occur.

The evidence for condition 1 was the consistently poor performance of participants in the assisted reflective
practice conditions relative to experiential practice, which contrasts with the good relative performance obtained in the group interaction/experiential practice condition of Experiment 1. Support for condition 2 was that the policies containing the most specific situations were associated with the best performance. Additionally, transfer performance was better in the specific hint condition relative to the general hint condition.

Explaining Reflective Abstraction Error

Reflective abstraction error is defined as the tendency to seek, find, and believe in simple rules or explanations of one's own task performance that may be grossly inadequate. Reflective abstraction error was manifested through three basic findings. First, reflecting within-task about one's model of the task (assisted reflective practice), or merely thinking ahead about the results of one's actions (prediction task), significantly reduced the degree of learning as indicated by task performance. Second, the mental models (written policies) of participants who reflected as they learned tended to be less valid and less specific than the mental models of experiential learners. Third, experiential and reflective knowledge tended to become dissociated when reflection occurred within-task and alternative ideas (other's policies in Experiment 1, or valid hints in Experiment 2) are provided to learners.

83
These findings are generally consistent with those reported in the literature on implicit learning of artificial grammars. Learners studying letter-strings generated by an artificial grammar under instruction to discover the rules of the grammar (reflective learning) tend to perform worse on a subsequent discrimination test than do learners instructed only to memorize the letter-strings (experiential learning). Rule-search instructions seem to cause learners to misclassify consistently some nongrammatical strings as grammatical (Brooks, 1978; Reber, 1976; Reber et al., 1980), indicating that they are following invalid rules.

Given that the rules of most artificial grammars are very complex and learners are given no feedback about the correctness of their developing rules, it is not surprising that they should fail to discover completely valid rules (Brehmer, 1980). However, the reflective practice procedure used in the current investigation was devised specifically to help learners discover the rules governing the sugar production task. They were allowed periodically to write down their rules. Additionally, they were given feedback about the overall quality of their written rules. Finally, and most important, they were required, under conditions of assisted reflective practice, to evaluate their rules by following those rules as they controlled sugar production. Incredibly, all of this support
contributed to poorer quality written policies which are evidence of less valid reflective knowledge. This support for reflection also contributed to poorer task performance.

Reflection abstraction error apparently can be prevented as demonstrated by Mathews et al. (1996). Contrary to the findings reported here, Mathews et al. (1996) found a large positive effect of reflective practice on both reflective knowledge (policy quality) and performance using the same sugar production task and a reflective practice condition very similar to that employed in this study. The reason for the very different results in the two studies is most likely to be found in differences in the reflective practice procedures used in the two studies.

In the Mathews et al. (1996) study, policy feedback comprised both a global score indicating overall policy quality and information concerning the relative effectiveness of individual policy statements or rules. In effect, the scored policies returned to the participants at the beginning of a session in Mathews et al. (1996) were somewhat like the hints given participants in Experiments 2 and 3 in the current investigation. Good rules were generally specific cases. Thus, the feedback pointed to these good specific rules that subjects included in their policies. Recall that the specific hints significantly improved performance and reflective knowledge.
The reflective practice procedure of Mathews et al. (1996) also instructed participants about which of their rules "best" applied in each situation during reflective practice. In a way, the simulation served somewhat as an expert coach guiding a student through the just the right actions at just the right time (Ericsson et al., 1993) during reflective practice.

In the current investigation, policies were returned to participants with a single score indicating the validity of the policy as a whole. Participants were not informed about the relative quality of individual rules or statements in their policies. Also, they were not guided as to which rule to apply during practice. The contrast in method and outcome between Mathews et al. (1996) and the experiments in the current investigation suggest that a key to thwarting reflective abstraction error is to keep reflection focused on specific experiences with the task. When learners are reflecting on their experiences with the sugar production task and not on their abstract theories or beliefs about the task, they are successful.

The performance of participants in the group interaction conditions of experiment 1 provides the strongest evidence for the importance of keeping reflection focused on task experiences. In those conditions, assisted reflective practice was most damaging; learners practicing reflectively performed much worse than those practicing
experientially. More important, there was a striking difference between the written policies of the two conditions. Policies of learners having experiential practice tended to be lists of highly valid and specific situations (e.g., "when production is at 3000 tons, always use 500 workers"). By contrast, policies of learners having assisted reflective practice contained few references to specific situations. The only difference in procedure between the two conditions was the mode of practice (reflective vs experiential). Therefore, it could be that practice mode was completely responsible for the differences in performance. However, the written policies were qualitatively different in the two conditions. Consequently, during the group discussion periods, discussion was focused on different kinds of knowledge. It is possible that the differences in performance were due to this difference in discussion content. When discussion was focused on specific or concrete situations (e.g., in the experiential condition) performance benefitted. By contrast, when discussions were focused on general rules, performance suffered. This possibility is supported by the finding of Experiment 2 in which performance under reflective practice conditions was good when learners were given hints containing specific cases. That is, reflecting on one's possibly flawed understanding of the task while one learns can have a detrimental impact on understanding.
This negative impact of reflective practice is contrary to what is typically found in studies of problem solving and transfer (Ahlum-Heath & DiVesta, 1986; Berry, 1983; Chi et al., 1989; Gagne & Smith, 1962; Wilder & Harvey, 1971). Good problem solvers are more likely to exhibit spontaneous use of metacognitive strategies in which they reflect on what they know or don't know and evaluate the success of their activities (Biemiller & Meichenbaum, 1992; Chi et al., 1987; Chi et al., 1994). Requiring solvers to justify or explain each step toward solution also has been found to facilitate problem solving (Berardi-Colletta et al., 1995; Chi et al., 1994).

An exception to this positive effect of reflective processing is found within a particular class of problems known as insight problems (Wertheimer, 1959). These are problems which have a high probability of leading to an impasse and whose solution is usually associated with an "ah ha" experience in which the solver suddenly obtains the solution (Schooler, Ohlsson, & Brooks, 1993). Schooler et al. (1993) found that a requirement to think aloud while working on insight problems reduced the likelihood of arriving at the correct solution. Their explanation of this result is that verbalization activates knowledge and processes that overshadow the nonverbal processes necessary for insight to occur.
Can "overshadowing" explain the negative effect of reflective practice? The main difficulty with this account is that the sugar production task does not have the central attribute of an insight problem. Insight problems are characterized by the sudden appearance of a solution. In the sugar production task insight should reveal itself in the form of discontinuous learning curves for individual learners. However, improvements in performance tend to be incremental for the sugar production task. Thus, insight does not describe the solution process in the sugar production task.

One might explain reflective abstraction error as the inevitable result of attempting to understand one's experience in the face of unreliable input. For example, it has been suggested that the poor quality of clinical judgment, as compared with other empirically based approaches to judgment, follows from a combination of the unrepresentativeness of the cases experienced, a lack of timely and accurate feedback about one's judgments, and biases people have which prevent them from properly evaluating the evidence of their experiences (Brehmer, 1980; Camerer & Johnson, 1991; Klayman & Ha, 1985; Tversky & Kahneman, 1973). That is, the less than stellar performance of human judges and decision makers is the best they can do in a noisy environment.
This diagnosis is most certainly correct. The problems of representativeness of experiences and paucity and reliability of feedback are evident (Brehmer, 1980). Also, people's biases have been well documented (Kahneman, Slovic, & Tversky, 1982; Klayman & Ha, 1985; Tversky & Kahneman, 1982). However, the implication of this diagnosis is that the poor performance of human judges is due to an adaptive agent (human learner) being faced with an environment that is ill-structured. Thus, the fact that the learner ends up not well adapted to the task environment should not be conceived as an error but as the best solution in the absence of good "data".

An important question to ask at this point is whether the sugar production task is just this sort of ill-structured task? The answer is, it is not. Feedback in the sugar production task is immediate and accurate, with a small exception made for the noise added to output. This suggests that reflective abstraction error is a response to something other than an ill-structured and noisy environment.

Perhaps the key to understanding the negative effect of reflective practice in the sugar production task is the nonlinear relation between number of workers used and sugar production. People tend to seek simple positive relationships between task variables (Hammond & Summers, 1965; Sanderson, 1989; Sniezek, 1986). Participants
reasonably expect that increasing the size of the workforce will always raise the level of production and similarly that using fewer workers will always lower the production level. That is, they assume a simple causative model in which each worker produces some fixed quantity of sugar (Sanderson, 1989). In fact, the relationship between production level and workforce size is more complex. The effect of changing the workforce will vary as a function of the production level achieved on the previous trial. An increase in the size of the workforce will sometimes raise production and sometimes lower production. As can be verified from the equation relating workers and production, production level will increase only when the number of workers selected is greater than one tenth of the current production level. For example, assume that on a trial the production level is 5000 tons and the current workforce size is 300. An increase of workers to 350 would lower production because 350 is less than one tenth of 5000. Assuming that 350 workers are selected, the resulting production would be 2000 (disregarding the random element). Then if workers are increased again, to 400 workers, production level will now increase because 400 is greater than one tenth of 2000.

Learners required to reflect as they practice are apparently unable to discard this common sense model of the sugar production task. Even when faced with feedback that
challenges the model (e.g., when an increase in workers leads to a decrease in production level) learners are apparently unable to change their representation of the task (e.g., Luchins, 1942). Perhaps the common sense model is protected from evidential challenges through the formation of exception rules that account for those situations in which the simple model fails (e.g., Holland et al., 1986). Or, similarly, perhaps the simple model persists through a self-perpetuating encoding process in which all ambiguous or contrary evidence is interpreted as being consistent with the model (Lewicki, Czyzewska, & Hill, 1997).

Regardless of whether learners come to the task with strong beliefs that conflict with the task or they adopt those beliefs in the absence of a better model, the conditions of assisted reflective practice may strongly activate one's mental model. Once activated, it influences one's interpretation of experience and thereby can lead to the abstraction of knowledge that is inconsistent with the task (e.g., Lewicki et al., 1997).

Feltovich, Spiro, and Coulson (1989) argue that people engage in a number of "conceptual biases" when faced with learning very complex material. These conceptual biases refer to tendencies to simplify complex concepts during acquisition. For example, when faced with comprehending a system with many interacting variables, one tends to focus
on one or two variables at a time and attempt to put the system back together after all of the parts are understood. Unfortunately, when interactions are present, this can lead to a misunderstanding of the complete system. Thus, Feltovich et al. argue, misconceptions arise as a natural result of attempting to understand a complex system. Further, the misconceptions arise out of actions that the learner takes to reduce the complexity.

It is not clear that this kind of analysis task complexity applies to the sugar production task (c.f., Hayes & Broadbent, 1988). Although some authors have referred to this task as a complex system (e.g., Broadbent et al., 1986), its complexity is not due to large numbers of interacting variables. Its workings are very simple and learners receive immediate and accurate feedback about their performance. Whatever complexity it exhibits is due to violations of learners' expectations in terms of the kinds of relations that they seek (e.g., Hammond & Summers, 1965; Sniezek, 1986). Thus, the difficulties that learners have are not due to any intrinsic complexity of the task, but rather to the "perceived" complexity and an incongruence between their expectations and the actual behavior of the system.

Implications

The results of this investigation suggest that the best way to support reflection for novices beginning to learn a
difficult task, and thereby prevent or reduce reflective abstraction error, is to direct their reflection toward the concrete data of experience with the task rather than the conjectures of their theories about the task. Experience, and our memories of it, are richer in detail than any theories of the phenomena producing the experience. Also, experiential knowledge does not seem to be so biased toward simple relations between variables. Thus, experiential practice followed by post-task discussion may provide the inductive base necessary to counter wayward theory (Brehmer, 1980; Mathews et al., 1989).

This prescription for combating reflective abstraction error is consistent with descriptions of the developmental process which one must follow to advance in many skills. For example, first language acquisition is certainly a case where substantial experience precedes the acquisition of valid reflective knowledge. The difficulties faced by adults when learning a second language may occur in part because they already have a language model (i.e., grammar) and the cognitive capacity to reflect on that model as they experience the new language (e.g., Johnson & Newport, 1989). Certainly adults have much more sophisticated theories about grammars than do children. Possibly, children with their more limited cognitive capacity are less able to reflect on the input.
In a similar vein, Bloom's (1985) discussion of exceptional performers in skills ranging from music to athletics suggests that the childhood experiences of all exceptional performers are characterized by play and exploration in their chosen skill for no other reason than their own enjoyment. Formal structured study typically comes much later in their development.

The importance of the theories learners bring, or don't bring, to any learning situation is most strongly apparent in the literature on misconceptions, or alternative conceptions (Wandersee et al., 1994), in science education. The difficulty students have in learning subjects such as simple mechanics (e.g., Caramazza, McCloskey, & Green, 1980; diSessa, 1982; McDermott, 1984), electricity (e.g., Cohen, Eylon, & Daniel, 1983; Heller, 1987), or the nature of matter (e.g., Novick & Nussbaum, 1978; Wandersee, 1983), has been found to be related to the conceptions they bring to the classroom and how those conceptions interact with the scientific concepts to be learned (Osborne, Bell, & Gilbert, 1983). The theories students bring to the classroom can lead to suboptimal learning outcomes.

One's natural response to the difficulties of these students is to challenge their concepts directly. However, the results of this investigation suggest that directly engaging people's reflective knowledge can lead to reflective abstraction error, unless their reflection is
directed toward discussion of experiential knowledge, or it is guided by an external source that can reinforce correct choice and application of reflective rules.

**Future Directions for Research**

The strong negative impact of reflective practice reported in these experiments requires replication. This is particularly important considering the contrast between the current findings and those of Mathews et al. (1996) in which reflective practice, combined with group discussion, led to superior performance. A logical follow-up would compare the two methods within a single experiment. An important variable to investigate in such a replication would be the type of feedback learners receive about their reflective knowledge (i.e., their written policies).

A possible limitation of this study is the reliance on the sugar production task. One could argue that Reflective Abstraction Error is limited to situations where there is an incongruence between the cover story describing the system (sugar production) and the actual behavior of the system. This limitation could be addressed in a couple of ways.

One way to address this limitation would be to investigate the effect of different cover stories (e.g., the "person control task", Stanley et al, 1989) varying in their degree of congruence with task behavior. Possibly,
Reflection Abstraction Error would be systematically related to this variable.

Another possibility would be to select less artificial tasks. Two such examples are the city transportation task of Broadbent et al. (1986) and the simulated economic system used by Broadbent and Aston (1978). Finding Reflective Abstraction Error in tasks such as these would support the generalization of these findings beyond the laboratory.

The strong negative effects of reflective practice reported in this investigation might challenge one’s faith in thoughtful reflection. However, we should find comfort in Schumacher’s observation that more experiential ways of knowing can guide us in our journey.

Yet a man who uses an imaginary map, thinking that it is a true one, is likely to be worse off than someone with no map at all; for he will fail to inquire whenever he can, to observe every detail on his way, and to search continuously with all his senses and all his intelligence for indications of where he should go. (E. F. Schumacher, Small is Beautiful)
REFERENCES


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Florida Artificial Intelligence Research Symposium.
111-115.

effects of verbalization on problem solving. Journal
of Experimental Psychology, 63, 12-18.

Gagne, R. M., & White, R. T. (1978). Memory structures and
learning outcomes. Review of Educational Research, 48,
187-222.

Gibson, J. J. (1979). The ecological approach to visual

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem-

dependence on linear and non-linear cues.
Psychological Review, 72, 215-224.

current electricity problems. In J. D. Novak (Ed.),
Proceedings of the Second International Seminar on
Misconceptions and Educational Strategies in Science
Department of Education, Cornell University.


the knowledge of experts from the perspective of
experimental psychology. The AI Magazine, 8, 53-66.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard,
P. R. (1986). Induction: Processes of inference,

Howard, J. H., & Ballas, J. A. (1980). Syntactic and
semantic factors in the classification of nonspeech
transient patterns. Perception and Psychophysics, 28,
431-439.

effects in second language learning: The influence of
maturational state on the acquisition of English as a

101

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


APPENDIX A

COMPUTER SIMULATION PROCEDURES

The primary purpose of the simulation program was to eliminate the subjectivity involved in selecting from a policy the most appropriate rule to apply on each trial. To that end, each policy was translated into a set of if-then rules (e.g., "if production is above target then use fewer workers"). The rules were submitted to a computer program that performed the same 40-trial test as performed by experimental participants.

Translation of Policies into Rules

Every attempt was made to translate every statement contained in each policy into a rule for selecting workforce size. Each rule was of the form: if <conditions> then <action>. Conditions could be any proposition about the state of the sugar production system. Conditions could refer to prior states (e.g., "if 400 workers were used on the previous two trials and production increased then..."). However, conditions were not allowed which referred to states that occurred in a prior block of ten trials. Because participants often stated rules with no condition (e.g., "use 600 workers"), empty conditions were allowed. For example, the rule above would be translated as "if <> then use 600 workers".

The actions of rules were allowed to be single values (e.g., use 350 workers) or lists or ranges of values.
This was necessary because policy statements often referred to workforce sizes in relative (e.g., "more") or abstract (e.g., "low") terms. When translating statements into rules, these cases were handled in the following way: (a) "more, or less, workers than X" was interpreted as a randomly selected value of workers between X and the maximum or minimum number of workers allowed, respectively; (b) "a high, or low, number of workers" was taken to mean a randomly selected number of workers above 750 or below 450 respectively; and (c) "an increasing, or decreasing, number of workers" was interpreted the same as in (a).

On occasion, participants would include mathematical formulas into the condition of a rule. Whenever possible, these formulas were translated into procedures that would check the state of the sugar production system to determine if the condition was satisfied. Similarly, the actions of rules sometimes were in the form of formulas (e.g., "if <action> then <take current production level and divide by 10 to get worker level>"). In this case, the action was translated into procedure that would output a list of one or more values for workforce size.

**Simulation Procedure**

The same 40-trial test performed by all participants was used as the simulation task. On each trial of the simulated task, all rules whose conditions are satisfied are first identified. Each rule then casts one vote for
each possible workforce size implicated by its action. For example, if the rule "if <production is low> then <use more than 600 workers>" had its condition satisfied, then it would cast one vote each for 650, 700, 750, 800,...1200 workers. The actual number of workers chosen on that trial is simply the number of workers receiving the most votes.

On trials where no rules applied (their conditions were not satisfied), the workforce size from the previous trial was repeated. When no rule applied on the first trial of a block, a random value was chosen for workforce size.
### APPENDIX B

**REPEATED-MEASURES ANALYSIS OF PERFORMANCE IN EXPERIMENT 1**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mode</td>
<td>1</td>
<td>1502.1</td>
<td>1502.1</td>
<td>10.90*</td>
</tr>
<tr>
<td>Exposure</td>
<td>2</td>
<td>4036.8</td>
<td>2.0</td>
<td>0.01</td>
</tr>
<tr>
<td>Mode X Exposure</td>
<td>2</td>
<td>402.8</td>
<td>201.4</td>
<td>1.46</td>
</tr>
<tr>
<td>Error(between)</td>
<td>109</td>
<td>15022.2</td>
<td>137.8</td>
<td></td>
</tr>
<tr>
<td><strong>Within subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Session</td>
<td>2</td>
<td>3112.4</td>
<td>1556.2</td>
<td>37.44*</td>
</tr>
<tr>
<td>Session X Mode</td>
<td>2</td>
<td>746.1</td>
<td>373.0</td>
<td>8.98*</td>
</tr>
<tr>
<td>Session X Exposure</td>
<td>4</td>
<td>91.9</td>
<td>23.0</td>
<td>0.55</td>
</tr>
<tr>
<td>Session X Mode X Exposure</td>
<td>4</td>
<td>275.3</td>
<td>68.8</td>
<td>1.66</td>
</tr>
<tr>
<td>Error(within)</td>
<td>218</td>
<td>9060.5</td>
<td>41.6</td>
<td></td>
</tr>
</tbody>
</table>

*p<.05.
APPENDIX C

ANALYSIS OF FINAL TEST PERFORMANCE IN EXPERIMENT 1

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode</td>
<td>1</td>
<td>1411.1</td>
<td>1411.1</td>
<td>16.05*</td>
</tr>
<tr>
<td>Exposure</td>
<td>2</td>
<td>33.1</td>
<td>16.5</td>
<td>0.19</td>
</tr>
<tr>
<td>Mode X Exposure</td>
<td>2</td>
<td>562.4</td>
<td>281.2</td>
<td>3.20*</td>
</tr>
<tr>
<td>Error</td>
<td>109</td>
<td>9580.4</td>
<td>87.9</td>
<td></td>
</tr>
</tbody>
</table>

*p<.05.
APPENDIX D

ANALYSIS OF FINAL POLICY QUALITY IN EXPERIMENT 1

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode</td>
<td>1</td>
<td>1067.4</td>
<td>1067.4</td>
<td>10.06*</td>
</tr>
<tr>
<td>Exposure</td>
<td>2</td>
<td>340.7</td>
<td>170.4</td>
<td>1.61</td>
</tr>
<tr>
<td>Mode X Exposure</td>
<td>2</td>
<td>1223.3</td>
<td>611.6</td>
<td>5.76*</td>
</tr>
<tr>
<td>Error</td>
<td>109</td>
<td>11569.9</td>
<td>106.1</td>
<td></td>
</tr>
</tbody>
</table>

*p<.05.
APPENDIX E

ANALYSIS OF CONTROL GROUPS IN EXPERIMENT 1

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>2</td>
<td>1269.1</td>
<td>634.6</td>
<td>6.04*</td>
</tr>
<tr>
<td>Error</td>
<td>57</td>
<td>5984.9</td>
<td>105.0</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Treatment groups contained in this analysis were: practice-only, predict-only, and no exposure/ experiential.

*p<.05.*
APPENDIX F

REPEATED-MEASURES ANALYSIS OF PERFORMANCE IN EXPERIMENT 2

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>2</td>
<td>3248.8</td>
<td>1624.4</td>
<td>8.91*</td>
</tr>
<tr>
<td>Error(between)</td>
<td>49</td>
<td>8928.9</td>
<td>182.2</td>
<td></td>
</tr>
<tr>
<td><strong>Within subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Session</td>
<td>2</td>
<td>7575.7</td>
<td>3787.8</td>
<td>93.34*</td>
</tr>
<tr>
<td>Session X Treatment</td>
<td>4</td>
<td>1364.7</td>
<td>341.2</td>
<td>8.41*</td>
</tr>
<tr>
<td>Error(within)</td>
<td>98</td>
<td>3977.1</td>
<td>40.6</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05.
### APPENDIX G

**ANALYSIS OF FINAL TEST PERFORMANCE IN EXPERIMENT 2**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>2</td>
<td>2019.3</td>
<td>1009.6</td>
<td>11.42*</td>
</tr>
<tr>
<td>Error</td>
<td>49</td>
<td>4331.7</td>
<td>88.4</td>
<td></td>
</tr>
</tbody>
</table>

*p<.05.*
APPENDIX H

ANALYSIS OF FINAL POLICY QUALITY IN EXPERIMENT 2

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>2</td>
<td>3022.6</td>
<td>1511.3</td>
<td>10.37*</td>
</tr>
<tr>
<td>Error</td>
<td>49</td>
<td>7143.0</td>
<td>145.8</td>
<td></td>
</tr>
</tbody>
</table>

*p<.05.
APPENDIX I

REPEATED-MEASURES ANALYSIS OF PERFORMANCE IN EXPERIMENT 3

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mode</td>
<td>1</td>
<td>34.4</td>
<td>34.4</td>
<td>0.25</td>
</tr>
<tr>
<td>Hint</td>
<td>2</td>
<td>1993.9</td>
<td>996.9</td>
<td>7.13*</td>
</tr>
<tr>
<td>Mode X Hint</td>
<td>2</td>
<td>113.9</td>
<td>56.9</td>
<td>0.41</td>
</tr>
<tr>
<td>Error (between)</td>
<td>115</td>
<td>16087.6</td>
<td>139.9</td>
<td></td>
</tr>
<tr>
<td><strong>Within subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Session</td>
<td>2</td>
<td>4820.9</td>
<td>2410.5</td>
<td>60.92*</td>
</tr>
<tr>
<td>Session X Mode</td>
<td>2</td>
<td>64.4</td>
<td>32.2</td>
<td>0.81</td>
</tr>
<tr>
<td>Session X Hint</td>
<td>4</td>
<td>907.7</td>
<td>226.9</td>
<td>5.73*</td>
</tr>
<tr>
<td>Session X Mode X Hint</td>
<td>4</td>
<td>119.2</td>
<td>29.8</td>
<td>0.75</td>
</tr>
<tr>
<td>Error (within)</td>
<td>230</td>
<td>9101.3</td>
<td>39.6</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05.
APPENDIX J

ANALYSIS OF FINAL TEST PERFORMANCE IN EXPERIMENT 3

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode</td>
<td>1</td>
<td>3.5</td>
<td>3.5</td>
<td>0.06</td>
</tr>
<tr>
<td>Hint</td>
<td>2</td>
<td>769.8</td>
<td>384.9</td>
<td>6.62*</td>
</tr>
<tr>
<td>Mode X Hint</td>
<td>2</td>
<td>6.9</td>
<td>3.5</td>
<td>0.06</td>
</tr>
<tr>
<td>Error</td>
<td>115</td>
<td>6685.2</td>
<td>58.1</td>
<td></td>
</tr>
</tbody>
</table>

*p<.05.
### APPENDIX K

#### ANALYSIS OF FINAL POLICY QUALITY IN EXPERIMENT 3

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode</td>
<td>1</td>
<td>34.3</td>
<td>34.3</td>
<td>0.22</td>
</tr>
<tr>
<td>Hint</td>
<td>2</td>
<td>85.5</td>
<td>42.7</td>
<td>0.27</td>
</tr>
<tr>
<td>Mode X Hint</td>
<td>2</td>
<td>259.0</td>
<td>129.5</td>
<td>0.82</td>
</tr>
<tr>
<td>Error</td>
<td>115</td>
<td>18101.3</td>
<td>157.4</td>
<td></td>
</tr>
</tbody>
</table>

*p<.05.
VITA

Lewis G. Roussel, Jr. is the youngest son of the late Lewis G. Roussel, Sr., and Barbara Francis Roussel. Born in Baton Rouge, Louisiana, on October 24, 1953, Lewis, Jr., received his undergraduate degree in Secondary Education from the University of New Orleans in 1977 and his master's degree in Psychology from Louisiana State University in 1990. He is a candidate for the Doctor of Philosophy in Psychology at Louisiana State University in December of 1998.
DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: Lewis G. Roussel, Jr.

Major Field: Psychology

Title of Dissertation: Facilitating Knowledge Integration and Flexibility: The Effects of Reflection and Exposure to Alternative Models

Approved:

[Signature]
Major Professor and Chairman

[Signature]
Dean of the Graduate School

EXAMINING COMMITTEE:

[Signature]
Karen A. Kemp

[Signature]
James H. Vandersee

[Signature]
Joe Kastellec

Date of Examination: 9/4/98