IMPROVING PROBLEM SOLVING EFFICIENCY: THE WHAT AND HOW OF CAUTION

By

MARTIN E. KNOWLES

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2008

To my wife, son, and parents, you have made me who I am and I live for you and because of you

ACKNOWLEDGMENTS

I thank my mom and dad, who provided me and my brothers with everything we ever needed. They sacrificed so that I could receive an excellent education and always supported me in everything I did. I thank my wife for putting up with me, I know it’s not an easy job. She has done a phenomenal job raising our son, and she gives me strength every day. I thank my advisor, mentor, and friend, Peter. He has provided me with many lessons in and outside of the classroom, and I have learned so much from him. Last, but definitely not least, I thank my committee, Drs. Abrams, Berg, Brenner, and Fischler. Their flexibility, unselfishness, and willingness to help are something I will never forget.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

IMPROVING PROBLEM SOLVING EFFICIENCY: THE WHAT AND HOW OF CAUTION

By

Martin E. Knowles

December 2008

Chair: W. Keith Berg
Major: Psychology

A three-stage framework was proposed describing the reduction of illegal moves in a problem solving situation. The Caution Induction Framework (CIF) proposed that consequences for making or avoiding illegal moves in the first stage of the framework influenced the problem solver to increase the amount of attention devoted to the problem in the second stage. This increase in attention subsequently increased checking or evaluation behavior in the third stage resulting in the reduction of illegal moves.

Three experiments were presented to test specific aspects of the framework. The first experiment tested whether a punishment was necessary to reduce illegal moves or if threat alone, without punishment, was sufficient to yield the same result. The second experiment tested whether an aversive stimulus was necessary or if a positive consequence would yield illegal move reductions. Specifically, participants in the critical condition in the second experiment received candy for avoiding illegal moves and for solving the problem efficiently. The third experiment tested whether an instruction to check each move for legality would result in illegal move reductions. In addition, participants in the critical condition in the third experiment were asked to perform this behavior aloud while their voice was recorded. The results from the three

experiments mostly supported the proposed framework of CIF; however, future research is required to further test the assumptions of the framework.

CHAPTER 1
INTRODUCTION

People solve problems every day in a variety of settings. Unfortunately, these problems are not often handled as well as one would like. A person's mistakes on a problem may waste both time and effort. An important goal for psychologists is therefore to (1) determine what difficulties problem-solvers may encounter when working on such problems, and (2) develop techniques for overcoming these difficulties. Kotovsky and colleagues extensively studied the factors that contribute to problem difficulty (Kotovsky, Hayes, & Simon, 1985; Kotovsky & Simon, 1990). However, less attention has been devoted to overcoming these difficulties and improving problem-solving performance.

What makes a task a problem is that the solution is not immediately available to the problem-solver. If the solution were immediately evident then it would not be a problem; it would be a task that needed to be completed, but not solved. The solution is unknown, so the problem solver must take action to find the solution and this requires that the problem solver interact with the problem physically and/or mentally. Since knowledge of the problem is limited, the problem solver will often make several incorrect moves. These incorrect moves may include *legal* moves that take the problem-solver down the wrong path (e.g., turning south down a street when the final destination is north does not violate any laws, but is incorrect because it takes us away from our goal). Incorrect moves may also be *illegal* because they violate one of the rules of the problem (e.g., turning the wrong way down a one-way street violates a rule because it breaks the law, even if it would bring us closer to our final destination).

An illegal move, as the author has referred to them, may carry different consequences depending on the problem. It may have severe consequences, such as the termination of the problem or – in the most extreme cases – even injury or death (e.g., running a red traffic light and causing an accident). However, the rule violation may have less severe consequences that result in a minor penalty (traffic fine) or a simple warning (verbal warning to stop at all red traffic lights in the future). This dissertation was focused on two goals – first, to understand the consequences that aid a problem-solver in reducing illegal moves, and second to understand how the problem solver's behavior was modified by those consequences.

The prevalence of illegal moves and potential theories of how and why illegal moves are made will be discussed first. Next a literature review providing insight into illegal move reductions will be reviewed. Then an in-depth overview of Knowles and Delaney's (2005) problem-solving research will be provided as it acted as a basis for much of the work that was presented here. This will be followed by an overview and in depth presentation of the CIF framework for the reductions of illegal moves proposed in this dissertation. Finally, this chapter will be completed with a brief preview of the three experiments that were conducted.

Illegal Move Commissions: Frequency and Proposed Causes

Illegal moves may often account for a significant portion of a problem-solving episode. Knowles and Delaney (2005) reported that upwards of 20% of the total moves in the hobbits and orcs problem were illegal and Jeffries, Polson, Razran, and Atwood (1977) reported illegal moves rate averages up to 32.8% on similar problems. Therefore, techniques that reduce illegal moves would likely reduce solution lengths and possibly the time required to complete the problem.

To develop techniques for reducing illegal moves it would first be beneficial to understand the reasons why problem-solvers select illegal moves. Research suggests three major causes for illegal moves. The first cause is related to understanding of the problem, the second to limitations on our mental resources, and the third is related to the specifics of people's move-selection heuristics. I will consider each in turn.

Proposed Causes: Understanding

One reason for the selection of illegal moves could be that the problem-solver does not understand the problem fully or the rules that result in an illegal move. Understanding the problem refers to knowledge of the goal of the task and how to reach that goal. Understanding the rules refers to knowledge of the parameters in which a problem-solver must work in to solve the problem. The rules are the restrictions set up by the environment or those explicitly stated. Kotovsky and Simon (1990) found illegal moves when the problem solver had trouble understanding the problem or more specifically trouble understanding what constituted an actual move. However, additional research has found that problem-solvers select illegal moves even when the problem and the rules of the problem are well understood (Jeffries et al., 1977; Zhang & Norman, 1994; Knowles & Delaney, 2005). People often make illegal moves, but the reasons why remain unclear.

Proposed Causes: Resource Limits

One major explanation for illegal moves was that problem-solvers have resource limitations. Exactly what these resource limitations are has not always been specified, but usually the term resources refers to working or short-term memory limits, attentional capacity, and thinking speed. Jeffries et al. (1977) proposed that illegal moves were selected due to resource limitations that prevented problem-solvers from correctly calculating future states or from checking moves for legality at all. Resource limitations have also been proposed for performance deficits, such as a lack of planning, in other tasks as well. Atwood and Polson (1976) attributed a lack of planning in water jugs problems to short-term memory limitations. A lack of planning was also obtained by O'Hara and Payne (1998) in research involving the 8-

puzzle. This lack of initial planning seems to support the claim that problem solvers do not have the resources to plan complete solutions.

Recent research, however, indicates that problem solvers often do have the resources necessary to plan complete solutions and to avoid illegal moves. Delaney, Ericsson, and Knowles (2004) found that when instructed, participants were able to plan complete solutions to difficult water jugs problems without any feedback. O'Hara and Payne (1998) also found that when the cost of making a move increased on the 8-puzzle, participants engaged in planning more frequently. Furthermore, Knowles and Delaney (2005) found that increasing the cost of selecting illegal moves subsequently reduced the number of illegal moves made, even though the resources necessary for calculating future states and for remembering to check each move for legality were not altered. These findings indicated that participants have the resources necessary to perform more efficiently on these tasks. It is still unclear why illegal moves are made if the resources are available to avoid them.

Proposed Causes: Move Selection

One additional area of previous research that has received attention in the problem-solving literature is modeling how a problem-solver selects each move. Understanding what was done by the problem-solver in selecting a move would be of great value to understanding how and why illegal moves are made. However, how a problem-solver assesses a move as acceptable or not is not entirely understood. Researchers have proposed models to explain how moves were evaluated and selected. One model that can be used to explain the evaluation process that problem-solvers may have engaged in on such tasks as the hobbits and orcs problem was proposed by Jeffries, Polson, Razran, and Atwood (1977). In this model the evaluation function:

$$e_i = aM_i + bC_i + cP_i \tag{1-1}$$

was used to determine if a move was acceptable and exceeded the required criterion. In Jeffries et al.'s evaluation function (Equation 1-1) e_i was the value of the evaluation function for state i, in which there were M_i hobbits, C_i orcs, and P_i hobbit-orc pairs on the right bank and a, b, and c were constant positive weighting factors.

This model consisted of a three-stage process and in the first stage the problem-solver evaluated potential moves according to a specific noticing order. Noticing order referred to the idea that the problem-solver would always consider specific combinations of travels first as these combinations would likely have a higher probability of making greater advances towards the final goal state. If a move was evaluated and had a higher value than the current state and had not been visited before then it would have been selected with a specific probability α^{i+1}, where i was the number of moves that had been considered and rejected during this episode. If the move was recognized as being previously visited then it was selected with probability β. If no move was selected in Stage 1 then the model attempted to find a move to a novel state in Stage 2. If all moves were identified as previously visited then the model attempted to find the optimal move in Stage 3 or it randomly selected a move in this stage.

Once a move was selected in one of the three stages it was then checked for legality with the model's illegal move filter. The resulting states of moves were checked for legality. A move was rejected if it was determined that the move was illegal and executed if it was determined that the move was legal. Jeffries et al.'s (1977) illegal move filter described the probability of checking a move for legality as fixed probability $\in_1$ and the probability of correctly rejecting an illegal move as fixed probability $\in_2$. Jeffries et al. also proposed that there were hard-to-detect and easy-to-detect illegal moves and that easy-to-detect illegal moves were always rejected. This was one of the first models the author found that specifically addressed how a problem-solver may identify and reject illegal moves.

Another model that was designed to provide an understanding of how moves were selected was Lovett's (1998) ACT-R model of choice. Lovett & Anderson (1996) have also used ACT-R to understand how the prior experiences of success and failure affected future decisions on the Building Sticks Task (BST). Lovett's model assumed that problem-solvers selected their next move based on the highest expected gain according to the following equation:

$$E = PG - C \qquad (1\text{-}2)$$

where E was the expected gain of the selected move, P was the estimated probability of achieving the production's goal, G was the value of the goal, and C was the estimated cost to be expended in reaching the goal.

Lovett and Anderson (1996) presented participants with the building sticks task (BST) where participants had to add and subtract three different size sticks to create a stick equal to a specified target length (similar to the water jugs task used by Luchins, 1942). In this task the participants solved several of these puzzles and the authors were particularly interested in participants' tendency to select one stick over the others as their first move and how previous successes and failures of beginning with different sticks affected this decision. Lovett's model of choice, as applied to BST, focused on a problem-solvers' ability to reflect on these past successes and failures of operators to calculate E from applying an operator. This model was designed for BST, but could be modified to model move selections on other tasks as well. Many other move selection models exist for various problems, but to review them all here would not be practical.

Next I will turn to a discussion of studies that attempted to reduce illegal move rates. The amount of research in the problem solving literature that specifically looked at reducing illegal moves during a problem solving episode was limited. However, there were several works that used the number of illegal moves committed as a dependent measure. Such research provides insight into how problem-solvers reacted to specific manipulations and allowed for speculation on how problem-solvers approached and solved problems.

Early Research

Two relevant studies which presented manipulations that reduced illegal moves provided insight into increasing problem solving efficiency. Zhang and Norman (1994) changed the problem constraints in isomorphs of the Tower of Hanoi problem from internal rules (verbal or written constraints that were to-be-remembered: e.g. the tall peg cannot be placed in the middle hole even though it fits) to external rules (constraints not explicitly stated because they were embedded/implied by the physical environment: e.g. the pen cannot be pushed through the eye of the needle). Not surprisingly, changing the problem so that the environment prevented participants from making illegal moves resulted in illegal move reductions. Kotovsky and Simon (1990) took a very challenging problem where the available move options were not easily understood and presented participants with all the legal options at every problem state. With the legal options available participants reduced the number of illegal moves they made on isomorphs of the Chinese Ring puzzle. These studies presented manipulations that reduced illegal moves and provided insight into problem solving performance. However, in both cases the problem was altered and could have potentially reduced the resources required to avoid illegal moves. Therefore it is difficult to determine whether either study lent support to the resource limitation hypothesis or not (Jeffries, Polson, Razran, & Atwood, 1977). In addition, changing the environment (making internal rules external) and presenting additional information (providing participants with all potential legal moves), respectively, to reduce resource requirements will not often be possible in the real-world and such manipulations may alter the problem itself. Ultimately the limitations of these works to address the illegal move reductions made these results difficult to apply to other problems in the laboratory or real world.

Although the occurrence of illegal moves accounted for a significant portion of total moves, in at least some problems, the reduction of illegal moves in problem-solving does not appear to be widely studied. This may have been due to the idea that the resource limitation hypothesis (illegal moves occurred because we have a limited number of resources available while working on a problem, thus illegal moves were inevitable) was so widely accepted and therefore there was little that could be done to improve illegal move commissions. Another possibility was that it may have been overlooked that illegal moves played such a significant role in problem solving episodes. A third possibility was that the link between illegal move reductions on laboratory controlled experiments and the real world application of these findings were not obvious and thus such research did not seem beneficial.

Knowles and Delaney (2005)

The most extensive research on illegal move reduction was conducted by Knowles and Delaney (2005). An overview of that paper was offered here for two main reasons; (1) they intentionally tested manipulations that reduced illegal moves and (2) their work acted as a basis for much of the work presented in this dissertation.

Specifically, Knowles and Delaney proposed a 3-stage framework in an attempt to understand where and how different manipulations affected a problem-solvers selection and rejection of an illegal move. They presented three experiments demonstrating ways to reduce illegal moves on isomorphs of the missionaries and cannibals or river crossing problems. They looked at what knowledge was transferred when solving the same problem twice and knowledge that was transferred to a novel isomorph. Finally, they introduced the idea that *caution* (this term is subsequently defined) was a result of their manipulations and that this caution subsequently led to the reduction of illegal moves.

Knowles and Delaney (2005) presented a hierarchical 3-stage framework in an attempt to understand how and where manipulations would affect a problem-solver in their selection and rejection of illegal moves. The first stage asked whether or not an illegal move came to mind (*Generation-Rate Hypothesis*). If an illegal move did not then a legal move would be selected. If an illegal move did come to mind then the problem-solver moved to the second stage which asked whether or not the problem-solver checked the rules to see if the move was legal (*Caution Hypothesis*). *Caution*, which described this stage, referred to the increasing likelihood that a candidate move would be checked for legality. If the rules were not checked then the illegal move was committed. If the rules were checked then the problem-solver moved to the third stage which asked whether the rules were correctly checked (*Rule-Verification Hypothesis*). If the rules were not checked correctly then an illegal move was committed. However, if the rules were correctly checked then the illegal move would have been rejected and avoided and another move would have been considered. This framework described illegal move rejections and allowed their experimental manipulations to address different stages/hypothesis of the framework (the Generation, Caution, Verification framework of Knowles and Delaney was referred to as *GCV* herein).

In Experiment 1, Knowles and Delaney (2005) punished participants for illegal moves on the hobbits and orcs version of the river crossing problem to assess participant's ability to avoid illegal moves. In addition, participants were instructed to think aloud while working on the problem in an attempt to discover when and how participants selected and rejected illegal moves. When compared to a control group Knowles and Delaney found that those participants that were punished made fewer illegal moves while the number of legal moves made did not differ. When compared to a silent control group the legal and illegal moves for participants that were instructed to think aloud while solving the problem did not differ. These findings indicated that participants were able to avoid illegal moves when their attention was directed to do so and that thinking allowed had minimal to no effect on problem solving performance.

Experiment 2 was similar to Experiment 1 in that participants solved the same problem and the experimental group was penalized for making illegal moves while the control group was not penalized. However, there were two main differences; (1) there was no think aloud condition, all participants solved the problems silently, (2) each participant solved the same problem twice where the second solution attempt was solved with no penalties regardless of the instructions on the first problem. The results replicated those of Experiment 1 in that the penalty group made fewer illegal moves while the number of legal moves did not differ on the first solution attempt. The control group made fewer illegal moves on the second solution compared to the first, but there was no change in illegal moves for the penalty group. One of the more surprising results was that the benefits from being penalized on the first problem were sustained and carried over to the second solution attempt as the illegal moves made did not differ between solution attempts. The lack of a decrease from the first solution attempt to the second may have indicated a floor affect. Having each participant solve the same problem twice allowed the authors to assess if there were any sustained benefits from solving the first problem under penalty instructions. This also allowed for the comparison of the change in legal and illegal moves with practice on the same problem.

Experiment 3 was similar to Experiment 2 except that the second solution was replaced with an isomorph of the first problem. This second problem had the same underlying structure as the first problem, but did not have any outwardly obvious similarities. As in Experiment 2 regardless of the instructions on the first solution attempt all participants solved the second problem without penalty. The results largely replicated those of the first two experiments where the number of legal moves did not significantly differ between the penalty and control groups, but the penalty group made fewer illegal moves compared to the control. The results revealed that those in the penalty group continued to make fewer illegal moves on the second solution even when they were presented with a novel isomorph and no penalty. In addition, practice on the first problem did not reduce legal or illegal moves on the second problem and indicated that little or no learning was transferred from one isomorph to the other. These findings indicated that something that was not problem specific, potentially some type of general problem solving knowledge, was learned and transferred to the second solution attempt in the penalty group.

Experiments 2 and 3 asked participants to solve two problems and provided information as to what was learned and transferred. In Experiment 2, participants solved the same problem twice and made fewer illegal moves on the second solution attempt. In the third experiment participants solved one problem and then solved a novel isomorph of that problem, but this did not significantly reduce illegal moves on the second solution attempt. These findings lend support to the idea that problem specific learning gained from practice, and not general problem solving knowledge, was transferred to the second solution in Experiment 2 as there was lack of transfer of learning to a novel isomorph from practice alone in Experiment 3. However, in Experiment 3 the illegal move reductions, which followed a penalty, were transferred to a novel isomorph indicating that some type of general problem solving learning was acquired and transferred to a novel problem as a result of being penalized. This general problem solving knowledge was of great interest to Knowles and Delaney.

One main focus of the work by Knowles and Delaney was on understanding how the penalty manipulation affected participants in illegal move reductions. The second stage of their framework gave rise to the *Caution Hypothesis*, which stated that over time a problem-solver would potentially be more likely to check candidate moves for legality, thus reducing the number of illegal moves committed. The authors stated that this hypothesis seemed the most intuitively valuable to test because it likely allowed for general learning of a tendency to check moves for legality that could potentially be transferred to a novel problem. Knowles and Delaney conducted a regression analysis on legal moves, illegal moves, and illegal moves considered (but not made). The analysis revealed that legal moves accounted for significant variance (the more legal moves made the more illegal moves made), lending support to the *Generation Rate Hypothesis*. The analysis also revealed that once legal moves were statistically controlled the number of illegal moves considered and correctly rejected were inversely related to the number of illegal moves made (the more illegal moves correctly rejected the fewer illegal moves were made), lending support to the *Caution Hypothesis*. Based on these results caution may have played a large role in increasing problem solving efficiency and received additional attention in this dissertation work.

In summary, Knowles and Delaney (2005) provided invaluable groundwork for the current work presented here. Their work demonstrated that illegal moves could be reduced during a problem solving episode when the cost of making a move was increased (penalty for making illegal moves). Their results also indicated that problem specific learning was transferred to the same problem (Experiment 2) and that general problem-solving learning gained through penalty enforcement was transferred to a novel isomorph (Experiment 3). The questions that this research led to, which were highlighted in the current work included: (1) could other aversive manipulations besides penalty lead to illegal move reductions, (2) was an aversive manipulation required for reductions or could positive manipulations yield the same results, (3) were participants checking moves for legality more often or was their accuracy of those they checked improving over time?

Caution Induction Framework (CIF (pronounced "chēf"))

Knowles and Delaney's (2005) GCV framework proposed that a punishment for illegal moves induced what they termed "*caution*," which was defined as an increased frequency of checking moves for legality. However, caution may be only one of the factors that explained why fewer illegal moves occurred after punishing them. I will therefore present the *Caution Induction Framework (CIF)* as a heuristic tool for laying out what behaviors might have contributed to the effects of punishment on illegal moves. The CIF framework will provide a basis for the empirical tests to follow, and lay the groundwork for a future program of research on illegal move reduction.

A brief overview of the CIF framework will be provided first to facilitate a deeper base understanding of the framework. The framework has three stages (**C**onsequence, **I**nvolvement, and **F**acilitation), as shown in Figure 1-1. The first stage (*consequence*) referred to the external manipulations that affected problem-solving performance. The second stage (*involvement*) referred to the result of the manipulations that occurred in the first stage. The assumption of the framework was that the consequences in Stage 1 resulted in increased attention and thus the problem-solver became increasingly involved in the problem. The third stage (*facilitation*) specified what behaviors were altered to decrease illegal move commissions and that these changes were a result of increased attention devoted to the problem in Stage 2.

The third stage of CIF begins with the indication that increased attention may have resulted in increases in either the frequency or precision of certain behaviors. That is, problem-solvers may have performed these behaviors more often or more accurately to reduce illegal moves, respectively. The two main behaviors specified in CIF are evaluation and checking. These are conceptually similar to the evaluation functions and illegal move filter from Jeffries et al. (1977). *Evaluation* refers to the consideration of moves to determine which move has the highest potential of assisting the problem-solver in advancing towards the final goal state. Evaluation occurs prior to the selection of a candidate move and once a move is selected as the best potential move it might or might not then be submitted to be checked using GCV. *Checking* refers to testing a potential move for legality after the move has been selected, but prior to being executed. The checking behavior used by problem-solvers incorporates the GCV framework from Knowles and Delaney (2005).

Stage 1 (Consequence)

In Stage 1 (the *consequence* stage), the experimenter provided additional instructions or consequences for performance on the problem. Previous research has demonstrated that a minor punishment (rating words for pleasantness for 30 s) occurring immediately after every illegal move subsequently reduced the rate of illegal moves (Knowles & Delaney, 2005). However, it may be possible to obtain similar effects with less intrusive methods such as a threat or reward. Threat and reward are included in CIF, but it was not yet known whether these factors had the same effect on illegal move commissions as punishment. For now, threat and reward are assumed to have the same effect as punishment, but this is further addressed after completion of the experiments. CIF makes the assumption that any consequence of illegal moves in Stage 1 act upon the problem-solver to influence Stage 2.

Stage 2 (Involvement)

In Stage 2 (the *involvement* phase), CIF assumes that problem-solvers have the ability to perform better than they do and this is a result of the problem-solver not becoming fully involved/engaged in the problem or the problem-solver not allotting enough attention to the problem. The amount of attentional resources possessed by a problem-solver and how these resources are distributed across the problem were of importance to understanding human problem-solving. However, understanding the amount of available attention and how it may be allotted to a problem remained outside the scope of this project. The assumption of CIF is that the attention in Stage 2 is increased and that this results in increased activity in Stage 3 and ultimately a reduction in illegal moves.

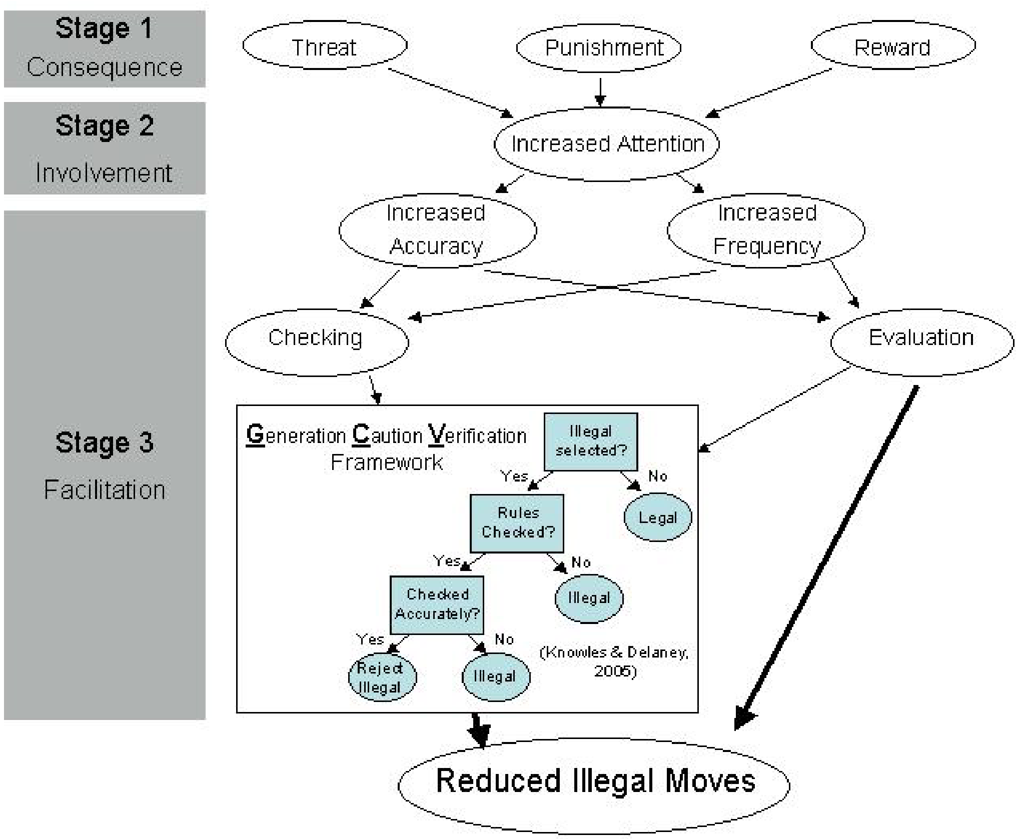

Figure 1-1. The three-stage Caution Induction Framework (CIF) demonstrating the process of reducing illegal moves in a problem solving episode.

Stage 3 (Facilitation)

Stage 3 (the *facilitation* phase) is a more complex stage compared to the rest of the framework and relies on increased attention passed on from Stage 2. Due to the complexity of Stage 3 the terms accuracy and frequency are first defined followed by subsections that explain checking and evaluation. The checking subsection discusses the Generation, Caution, Verification Framework from Knowles and Delaney (2005) and how increases in accuracy and/or frequency may affect problem-solving performance. The evaluation subsection does not focus on specific evaluation functions used by problem-solvers to reduce illegal moves-- it only assumes that evaluation of moves occurs and that increasing the frequency and quality of these evaluations indirectly reduces illegal moves. If people select moves randomly (with minimal evaluation), then one might expect a greater frequency of illegal moves compared to people who carefully consider what moves to select.

With the increased attention passed on from Stage 2 the problem-solver either increases the accuracy or the frequency of either checking or evaluation behaviors, which reduce illegal moves. An *increase in accuracy* means that the problem-solver performed this behavior with fewer mistakes compared to previous attempts resulting in an increased likelihood of not committing an illegal move. For example, if you check an illegal move for legality ten times, but only correctly reject the illegal move three times (a rate of 30%) then increasing correct rejections to five out of the ten (50%) would increase the number of illegal moves rejected from three to five. An *increase in frequency* simply means that the problem-solver performed this behavior more often. For example, if you only check three out of every ten moves for legality -- and you always correctly reject illegal checked moves -- then increasing frequency to checking five out of every ten moves would increase the number of correct rejections from three to five.

Either checking or evaluation behaviors can be increased in their accuracy or their frequency, resulting in fewer illegal moves. It is also possible that any combination of increasing the accuracy and/or frequency of either checking and/or evaluation simultaneously would result in reducing illegal moves.

Checking

Checking refers to a problem-solvers' verifying whether the resulting state of a potential move is valid or invalid according to the rules of the problem. If checking never occurs then the consideration and selection of an illegal move would likely always result in the commission of an illegal move. Knowles and Delaney (2005) provided us with a framework addressing how an illegal move was correctly rejected. To correctly reject an illegal move there were three steps that must have occurred according to Knowles and Delaney. An illegal move must have come to mind and been selected as a potential move, the problem-solver must have remembered to check the move for legality, and the problem-solver must have accurately checked the resulting state of the move and correctly rejected the illegal move. This framework, which is referred to here as the Generation, Caution, Verification (GCV) framework, was taken from Knowles and Delaney (2005) and has been applied to CIF in an attempt to understand how illegal moves were reduced following a penalty or another type of consequence (GCV is displayed in the box under the "Checking" oval in CIF in Figure 1).

Illegal moves can be reduced if the problem-solver becomes more accurate during the checking process. This increase in accuracy would have its effect in the third colored box of GCV. Even though a problem-solver checks a move for legality that does not guarantee that the move will be correctly rejected. The problem-solver may miscalculate the future state and erroneously think that the move is legal. Jeffries et al. (1977) proposed that calculation errors were due to resource limitations. However, the author believed that this was more likely due to insufficient resources being allotted to the process. With increased attention the problem-solver may allot additional resources to the checking process to correctly calculate the resulting state, resulting in more correct rejections of illegal moves.

Problem-solvers may not only become more accurate at checking moves, but they may check moves for legality more frequently. This increase in the frequency of checking would have its effect in the second colored box of GCV. With minimal attention allotted to the problem, problem-solvers may not bother to check the legality of moves at all. They may select a move and then execute it because there is little cost if the move proved to be illegal. However, when the cost of making illegal moves increases the problem-solver may attempt to verify the legality of every move before it is executed. Even if the accuracy of checking does not increase the number of illegal moves should still be reduced by the increase in the frequency of checking.

Evaluation

Stage 3 of CIF indicates that illegal moves could be reduced through evaluation; however, the current work does not explore evaluation and made no assumptions as to how a problem-solver evaluate moves. Determining the specific evaluation method used seems premature at this point, but it has been included for framework completeness. Evaluation refers to problem-solvers' assessment of a move to determine if the move they have selected reaches a selection criterion specified by the problem-solver. How the criterion is set and how each move is calculated is not fully known and is discussed in more detail below. Previously, in the Proposed Causes: Move Selection section a simple move evaluation model was described (Jeffries et al., 1977). CIF did not currently make any specific assumptions or proposals, other than those previously discussed in the GCV framework or those otherwise explicitly stated herein, as to how moves are evaluated. It assumes only that evaluation occurs and that the problem-solver's goal is to select the best candidate move that will advance towards the goal. If the criterion is met then the move is considered to be acceptable. If more than one of the considered moves meets or exceeds the criterion then the potential move that exceeded the criterion by the greatest margin is selected. The selected move is then either executed or checked for legality prior to being committed.

As with checking, illegal move reductions could potentially occur through either increasing the frequency or the accuracy of evaluation or both. With minimal attention devoted to the problem, problem-solvers may select moves without first evaluating them. This would likely result in the selection of several poor moves, which should be rejected, but are executed. *Poor moves* are defined as moves that do not bring the problem-solver closer to the goal state and this includes moves that are illegal because they would never advance the problem-solver towards the goal state. However, several illegal moves in the problem space may “appear” to advance the problem-solver closer to the goal; so would increased frequency of evaluation result in more illegal moves? This is a fair point and describes a real possibility; however without a specific evaluation function being used it is not possible to determine the probability of this occurring. However, although a specific move evaluation function is not assumed it would likely be the case that prior experience would affect subsequent moves. Prior experience would allow the problem-solver to reject illegal moves based on the premise that they have been previously committed or previously considered and correctly rejected. Therefore, with an increase in the frequency of evaluation the problem-solver would likely learn from previous mistakes and avoid additional illegal moves.

Problem-solvers may not only evaluate more frequently, but they may increase their accuracy at evaluation. To increase the accuracy of evaluation the problem-solver may increase the criterion a potential move must meet before it is selected. With a more stringent assessment of candidate moves the problem-solver would likely reduce not only the rate of illegal moves, but legal moves as well. One way illegal moves could be reduced would be if the increased criterion required that a move be legal. Such a requirement would likely involve the checking of candidate moves for legality as part of the evaluation process. Legal moves would be reduced because the increased criterion should avoid, in most cases, revisiting previously visited states and legal moves that take the problem-solver away from the goal. CIF's evaluation portion is not specifically tested therefore these claims can not be supported and additional research is required to determine the true effect of increasing the frequency and/or accuracy of evaluation on such problems.

Stage 3 of CIF received the most attention as it represents the place in the problem-solving episode where illegal moves are rejected. As is displayed in the framework, these rejections could occur through several different behaviors. The two primary behaviors are checking and evaluation and increasing the accuracy or frequency of either of these behaviors would likely result in illegal move rejections. In addition, increases in evaluation could likely also influence checking behaviors as they may be engaged in following the evaluation of a candidate move. Although Stage 3 received the most attention and appears to be the most important it also proved to be the most difficult to experimentally test as it is based upon untested assumptions of human problem-solving.

CIF Summary

In summary, CIF is a 3-stage framework that attempts to provide a better understanding of how consequences influence behavior to reduce illegal moves and increase problem solving performance. In Stage 1 the framework is restricted to only a few potential testable manipulations that may help to reduce illegal moves. Stage 2 is built upon the previous findings that problem-solvers may have greater abilities and additional resources to avoid illegal moves and plan ahead (Knowles & Delaney, 2005; Delaney, Ericsson, & Knowles, 2004; O'Hara & Payne, 1998). The third stage looks at two primary behaviors that would likely result in illegal move reductions, checking and evaluation. In addition, the third stage built upon the previous work of Knowles and Delaney and incorporated their GCV framework.

According to CIF the evaluation of a move may involve checking, but checking is not required for evaluation to result in a reduction in illegal moves and checking may occur and reduce illegal moves independent of evaluation. If a participant evaluates a move, but does not check the move for legality they may still benefit from the evaluation and select better moves that are more likely legal, resulting in the reduction of illegal moves. A participant may also select a move without evaluating that move, but still check the legality of the move and increase the probability of identifying the move as illegal, ultimately rejecting that move.

In conclusion, there are four questions of importance that needed to be addressed according to CIF: (1) if threat and reward yield the same result as punishment in Stage 1 (reduced illegal moves) (2) the amount of attention a problem-solver devotes to the problem after punishment, threat, and reward in Stage 2 (also how this attention may be devoted to different areas of the framework) (3) what actual behavior(s) problem-solvers engage in to reduce illegal moves in Stage 3 and (4) determining the evaluation process a problem-solver engages in when considering a move in Stage 3. This project attempted to answer the first question to determine if less aversive, less intrusive techniques could be substituted for punishment and yield similar results (Experiments 1 & 2). The finding that threat and reward have the same effect as punishment would support the idea that an increase in attention and not avoidance behavior was responsible for illegal move reduction. The second question was outside the scope of this project and should only be considered after it is determined that an increase in attention occurred after the manipulations in Stage 1. Attempting to determine the amount of attention devoted to the problem without first verifying that increased attention was responsible for the change in behavior would have been pointless. The third question was also addressed in this project to determine what behavior(s) problem-solvers engaged in after the manipulations in Stage 1 of CIF (Experiment 3). Understanding how problem-solvers were able to reduce illegal moves would enable us to discover better ways of increasing this behavior to improve problem solving efficiency. The fourth question was also outside the scope of this project and should only be considered after it is determined that problem-solvers engaged in evaluation processes. Attempting to determine the evaluation process used to select moves without verifying that an evaluation process took place would have also been pointless.

Experiments

Three experiments were presented in an attempt to obtain a deeper understanding of the role of caution in a problem-solving episode and how it changed based on the interaction between the problem solver and the problem. These experiments aided in determining what was required to induce caution and also helped determine what behavior was observed as a result of being cautious. The first experiment examined whether a threat of punishment without any punishment was sufficient to induce caution. The second experiment examined whether an instruction to be cautious or a reward for being cautious without any penalty or threat of penalty was sufficient to induce caution. Finally, the third experiment examined how the frequency and accuracy of legality checking was altered as a result of increased cautiousness.

CHAPTER 2
EXPERIMENT 1

Since illegal moves have accounted for a significant portion of total moves in at least some problems, knowing and understanding ways to reduce such moves would offer great benefits. In previous studies participants reduced the number of illegal moves they made after being penalized for illegal moves (Knowles & Delaney, 2005). However, these findings inspired the question of what other manipulations may have also reduced illegal moves. In particular, what other less aversive, less intrusive manipulations could have been implemented that would have reduced illegal moves? Less aversive manipulations would likely minimize unnecessary effects on the problem-solver and less intrusive manipulations would likely be easier to implement due to minimizing interference. Such advantages would potentially have real world applications for improving performance in business climates, the military, school, etc.

In the first experiment, participants in the two critical conditions were threatened that upon completion of the problem they would be penalized for the illegal moves they committed during their solution. One of the groups experienced the punishment prior to beginning the problem while the other group was simply told of the penalty without experiencing the punishment. Results from these two groups were compared to two additional groups, one that was penalized after every illegal move and one that was not penalized or threatened for making illegal moves. In addition, participants in all four conditions solved a second problem that had no penalty or threat of penalty to determine what learned information was transferred to a novel isomorph.

Threat, punishment, and reward were manipulations made on the problem and this was accounted for in Stage 1 of CIF (Figure 1-1). These manipulations also had a direct effect on Stage 2 of the framework where they increased the amount of attention devoted to solving the problem to deal with the so-called threat. People could have often been lackadaisical when they approached a problem that had little or no consequence for performance that was not optimal. This could have been thought of as having the motivation to do well being turned “off”. However, increasing the consequence or perceived consequence, as with threat, may have increased the amount of attention a person decided to devote to solving the problem as was shown in Stage 2 of CIF and this could have been thought of as turning motivation to the “on” position. This increase in attention or turning “on” motivation would have in turn had an effect on Stage 3 of the framework and would have ultimately resulted in the reduction of illegal moves.

The author believed that if the threat was perceived as real it should have been enough to influence the participant to engage in the problem more fully and should have resulted in an increase in behavior that would have reduced illegal moves. How the different manipulations of threat were perceived would determine how much attention was devoted to the problem and to what degree illegal moves were reduced. If the participant was told the threat, but they did not experience it before beginning the problem, they may have overestimated or underestimated the severity of the threat. If the threat was overestimated then the effect in Stage 2 would have been maximized and the participant may have reduced illegal moves to the degree that they were reduced in the penalty condition or even to a greater degree. If the threat was underestimated then the effect in Stage 2 would have been minimal and the reduction of illegal moves would have been less than that in the penalty condition and may have been similar to the no-cost control condition. Experiencing the threat before beginning the problem could have also displayed similar effects compared to the threat group that did not experience the penalty. After completing the initial penalty before beginning the problem the participant may have not perceived the penalty as very costly and this would have likely minimized the effect on Stage 2 (little or no increase in attention). In contrast, the participant may have perceived the penalty as very costly and this likely would have maximized the effect on Stage 2 (large increase in attention).

The author predicted that a participant would have been accurate at predicting the cost of the threat even when the threat was not experienced prior to beginning the problem. This would have resulted in illegal move reductions that were greater than the no penalty condition and similar to those of the penalty condition. However, it seemed likely that after performing the penalty, but prior to beginning the problem, the participant may have become even more attentive and may have showed illegal move reductions that were greater than those obtained in the no experience threat condition and closer to those in the penalty condition. If illegal move reductions for either of the critical conditions were greater than those in the no-cost condition then this would have provided evidence supporting the claim that less intrusive methods were capable of inducing caution.

Methods

Participants

Participants were recruited from General Psychology courses through the human subject pool at the University of Florida. Participation was voluntary and each participant received course credit for their participation. The duration of the experiment was approximately one hour and each participant was awarded two experimental credits. One hundred participants participated in this experiment; they were randomly assigned to one of the four groups. Twenty participants were unable to solve both problems correctly within the time limit and their data was therefore dropped from the analysis. Participants that were unable to solve both problems correctly in the time allowed were replaced until there were 20 participants in each of the four groups.

Problems

Hobbits & orcs

The hobbits & orcs problem (which was also referred to as the *missionaries and cannibals* or *river-crossing problem*) consisted of one boat and six travelers, three of which were hobbits and three of which were orcs. All travelers began on the left bank of a river, which was near the middle of the computer screen, and the boat began on the left bank at the bottom. The goal was to move all six travelers to the right bank of the river using the boat. However, the rules stated that the boat could only hold a maximum of two travelers at one time, and at least one traveler was required in the boat for it to cross the river. The rules also stated that at no time could the orcs outnumber the hobbits on either bank of the river because the orcs would then kill the hobbits. A button was located on the display to reveal the rules. If at any time a participant forgot one of the rules they could click on the button to look up any of the three rules. An example of the display seen by participants for this problem was presented in Figure 2-1.

The problem was written in Microsoft Visual Basic and presented on a Gateway desktop computer. Participants used the mouse to select icons representing the travelers to be moved and then selected the boat to send the selected travelers to the other bank of the river. After a traveler was selected it appeared at the bottom of the screen next to the boat. Clicking on a traveler a second time removed him from the boat and placed him back on the bank in the middle of the screen. If the participant added too many travelers to the boat, allowed the orcs to outnumber the hobbits, attempted to move the boat with no travelers selected, or violate the rules in any other way he/she was then notified via a message box and the move did not occur (problem description taken from Knowles and Delaney, 2005).

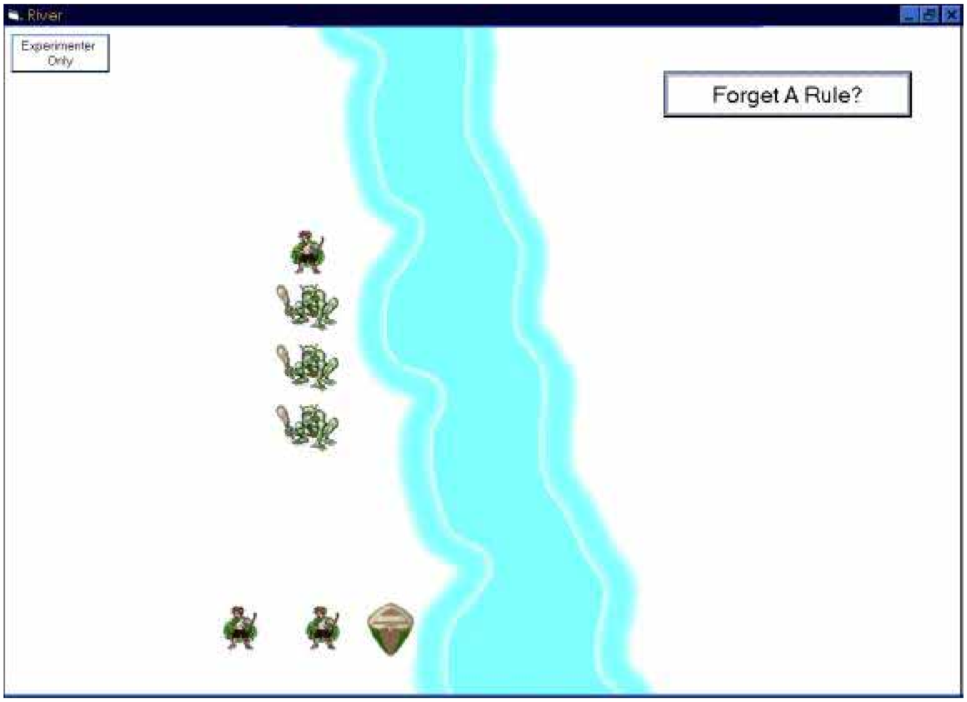

Figure 2-1. The display seen by the participant for the hobbits and orcs problem where two hobbits had been selected in the beginning state.

Titration

The titration problem was an isomorph of the hobbits and orcs problem in which the participant was asked to remove unstable isotopes from a white beaker. The beaker contained six isotopes, three blue and three orange. The goal of the problem was to remove all six isotopes from the beaker using a dropper to extract them. The participant began by removing isotopes and then alternated between adding and removing isotopes thereafter. However, the rules stated that the dropper would hold only two isotopes at one time, and at least one isotope was required in the dropper for the participant to add or remove isotopes. The rules also stated that if the number of blue isotopes in the white beaker did not equal the number of orange isotopes in the white beaker then the number of blue isotopes had to equal either zero (the minimum number possible) or three (the maximum number possible) or the isotopes would become unstable and explode. If at any time a participant forgot one of the rules, he or she could click on the button to display three additional buttons, one for each of the three rules.

The problem was written in Microsoft Visual Basic and presented on a Gateway desktop. A red “Remove” or “Add” button appeared on the screen indicating to the participant whether he/she was to be removing or adding isotopes to or from the white beaker at each step. When the “Remove” button was present the participant would use the computer mouse to click on the isotopes in the white beaker, which would then appear in the dropper. Clicking on the “Remove” button completed the move and made the “Remove” button disappear. The “Add” button then appeared along with a large blue and orange beaker. Clicking the blue or orange beaker added the appropriate color isotope to the dropper. Clicking the “Add” button completed the move and emptied the isotopes from the dropper into the white beaker. The “Add” button and the two colored beakers disappeared, and the “Remove” button appeared again.

The rules stated that it was impossible to have more than three blue or orange isotopes on the screen at any time and this included the contents of the white beaker, the dropper and the colored beakers. If a participant added too many isotopes to the dropper, attempted to remove or add isotopes with no isotopes in the dropper, allowed the isotopes to become unstable, or violated the rules in any other way he/she was notified via a message box, and the move did not occur. An example of the display seen by participants for this problem was presented in Figure 2-2.

Design

Four groups of participants solved two problems. The major difference between the groups occurred on the first problem when participants were instructed as to the consequences of making an illegal move. Different instructions were provided to each group informing them of the consequences of making illegal moves on the first problem. However, for the second problem all four groups received the same instructions – specifically, that there were no consequences of making illegal moves. The second problem was included to examine transfer effects. The four conditions differed on the first problem as follows:

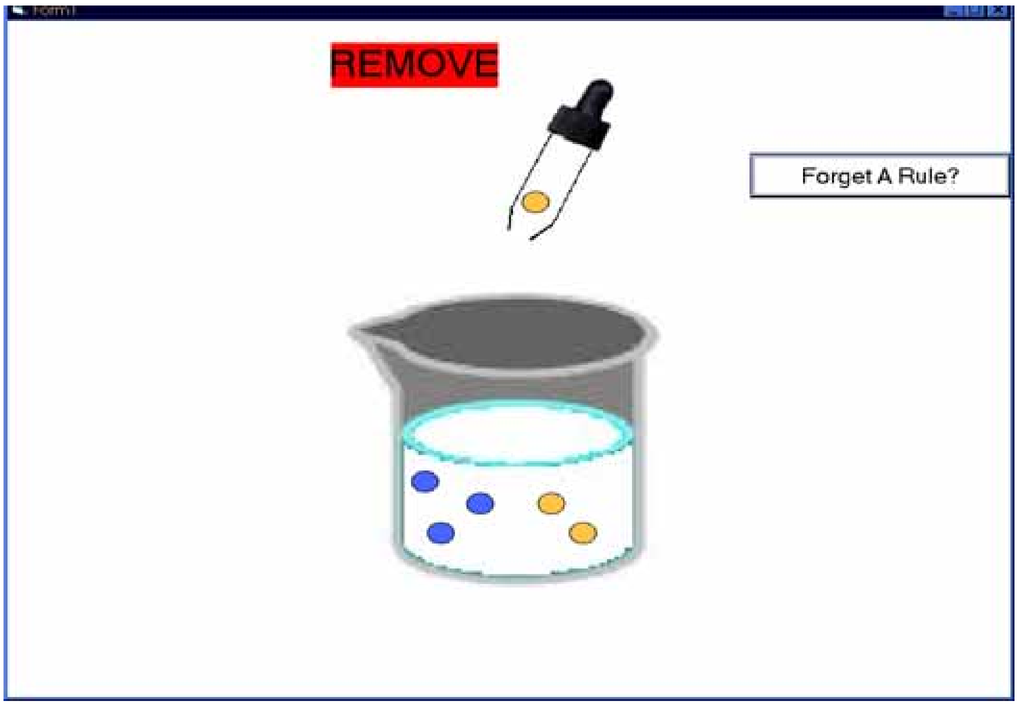

Figure 2-2. The display seen by the participant in the titration problem where one orange isotope had been selected in the beginning state.

In the *no-cost* group, a message box appeared after each illegal move indicating that an illegal move was made. The participant clicked an "OK" button to continue working on the problem from the last legal state. A participant was not penalized or threatened in any way. This group was identical to the no-cost group used in Knowles and Delaney (2005).

In the *punishment* group, a participant was informed before beginning the first problem that after each illegal move the problem would be paused and they would have to rate the pleasantness of words for 45 s. After every illegal move the participant was informed of the illegal move and the screen then turned gray. Words appeared in the center of the screen and the participant had to rate the words for pleasantness on a scale from 1 to 5 (1 being very unpleasant, 3 neutral and 5 very pleasant). After 45 s the screen returned to the problem and the participant was allowed to continue working on the problem from the last legal state they had reached. This group was identical to the punishment group used in Knowles and Delaney (2005).

In the *threat* group, a participant was instructed prior to beginning the first problem that for every illegal move he/she made the penalty of rating words for pleasantness would accumulate and he/she would have to complete the penalty after solving the problem. That is, for every illegal move the participant would have to complete 45 s of rating words for pleasantness. A participant in this condition was not penalized for making illegal moves and did not receive any additional information after making an illegal move. After making an illegal move a message box appeared stating that an illegal move was made and the participant was instructed to click "OK" and was then allowed to continue working on the problem from the last legal state.

Finally, the *experience* condition was identical to the threat condition except that prior to beginning the problem the participant completed the 45 s penalty phase once.

Procedure

Participants were tested individually and were randomly assigned to one of the four conditions. Half the participants solved the hobbits and orcs problem followed by the titration and the other half first solved the titration problem followed by the hobbits and orcs. To begin, the participant read the cover story for the first problem aloud and then read the three problem rules. The participant was asked to review the cover story and the rules and was instructed to study and memorize the three rules. Once the participant was able to recite the three problem rules to the experimenter without error the tutorial phase began.

During the tutorial phase, the participant was shown an example problem on the computer with instructions explaining how the problem worked and how he/she was to navigate the problem space. The participant was instructed to click on the "Forget a Rule?" button and then to click on each individual rule button and asked to recite each rule aloud. The participant was informed that the “Forget a Rule?” button was on the screen while he/she worked on the problems so that the participant could check the rules at anytime if he/she so desire. Next, the participant was instructed to move one of each character to the next position and then he/she was instructed to move both characters back to the starting place so that the participant could understand how the characters were moved. Then, the participant received instructions on violating each of the three rules in succession and was asked to violate each rule within the display to ensure that he/she fully understood what constituted an illegal move. The experimenter checked to make sure that the participant had done this correctly.

Finally, the participant was instructed to make three legal moves on the practice problem. The practice problem consisted of two hobbits and two orcs or two blue and two orange isotopes (instead of three of each as in the experimental problem), depending on which problem he/she solved first. At the end of the tutorial phase the participant was given the opportunity to ask questions that he/she may have had. The main goal of the tutorial phase was to ensure that the participant understood the problem, including the rules, the goal, and how to move the travelers.

The participant in the punishment condition was notified after the tutorial phase of the penalty for violating Rule 3 and received instructions on how to complete the word-rating task. The participant was penalized for Rule 3 violations only, there were no consequences for violations of Rules 1 or 2. Participants in both the threat and experience conditions were informed that they would have to complete the punishment upon completion of the problem, once for every illegal move. Those in the experience condition completed a sample of the punishment for 45 s before beginning the problem. In all conditions, testing began when the participant clicked on the mouse to initiate the program. If a participant was unable to solve the problem within the 20-minute time limit, that participant was then assisted in finishing the problem and his or her data was not included in the analyses. The maximum solution time of 20 minutes was chosen to restrict the session length to one hour.

After completion of the first problem the participant was asked to judge how many illegal moves he/she believed he/she had committed on the completed problem. These estimates were used to calculate a difference score between the participant's estimation and the actual number of illegal moves. The difference scores were compared between the groups to determine if accuracy differed. Participants in the threat and experience conditions were also informed that they would not have to complete the penalty and would be asked if they believed that they were going to receive the delayed punishment for making illegal moves. This question was asked to determine if participants believed the instructions of a delayed punishment. The participant was then instructed that he/she would now attempt to solve a different problem. There was no indication or mention during the experiment that the two problems were similar in any way. The participant was given a cover story for the second problem where he/she learned the rules and completed a tutorial phase, just as the participant did for the first problem. Before beginning the second problem all participants were instructed that there would be no penalty for illegal moves on the second problem. If a participant was unable to solve the second problem within the 20-minute time limit, that participant was then assisted in finishing the problem and his or her data from both problems was not included in the analyses.

Results and Discussion

Move data. For this experiment a 4 x 2 x 2 mixed factorial Analysis of Variance (ANOVA) was conducted: Group (No-Cost vs. Punishment vs. Threat vs. Experience) x Order (Hobbits first vs. Titration first) x Solution (First vs. Second). Group and order were between-subjects factors and solution was a within-subjects factor. Illegal moves committed on the first and second problem were of primary importance; however the number of legal moves and average time to complete each move were also analyzed.

Based on previous findings obtained by Knowles and Delaney (2005) no main effect of order, no main effect of solution, and no interaction for any of the factors for illegal moves, legal moves, and time were predicted. This would support the assumption that the isomorphs did not differ in difficulty, the order of presentation did not influence the results, and the results for the first solution did not differ significantly from the second solution. It was possible that a participant could transfer some general learning to the second problem and show a main effect of solution with improved performance on the second solution, but such a result was not found in Knowles and Delaney so such a result was not expected here either. It was also predicted that legal moves and total time would not differ significantly between the four groups. However, it was anticipated that a main effect of group would be obtained for illegal moves. A 4 x 2 x 2 mixed factorial ANOVA was conducted on legal moves made, average time to complete a move, and illegal moves committed with Group and Order as between-subjects factors and Solution as a within subjects factor.

Legal move data. The results for legal moves made revealed no main effect of solution, $F < 1$, indicating that the number of legal moves on the first and second problem did not differ statistically. There was no main effect of order, $F(1,72) = 1.27$, $p > .05$, $MSE = 244.42$, indicating that the number of legal moves made did not statistically differ based on which problem the participant solved first. There was also no main effect of group, $F < 1$, indicating that the number of legal moves made between the groups was not statistically different. These results lend support to the prediction that legal moves remain relatively unaffected by the manipulations made following illegal moves. Legal move averages and standard deviations for each group by solution were presented in Table 2-1.

Table 2-1. Legal moves made on the first and second solutions were displayed by group and solution and included mean and standard deviation (SD). All independent and pairwise comparisons were not statistically significant between groups.

	Group	Mean	SD
Solution 1	No-Cost	25.50	12.60
	Threat	20.85	11.43
	Experience	24.30	20.98
	Punishment	22.40	8.63
Solution 2	No-Cost	23.05	22.56
	Threat	20.70	10.04
	Experience	25.50	16.95
	Punishment	18.35	6.65

Move time data. The total time to complete the problem was divided by the total number of moves for both the first and second solution for each participant in order to derive average time per move in milliseconds (ms). The timer for the program that presented the problems began in the experience group when participants first experienced the penalty and did not pause or stop during the penalty in the punishment group. As a result of these continuous timers adjustments were necessary to account for the timer being engaged while the participant was not working on the problem.

The presentation of words in the penalty phase took 45 s -- that is, 3 s for each of the 15 words displayed. To account for the continuous timer 55 s was deducted from the total time to complete the first problem in the experience group prior to calculating move time averages. The 55 s includes 45 s for the presentation of the words during the penalty phase and 10 s for the brief instructions presented prior to the words during the penalty phase. In addition, 55 s was deducted from the total time for the first problem in the punishment group for the first illegal move and 45 s for each additional move because it was not required that the instructions be repeated after they were initially understood.

The results revealed no main effect for order $F < 1$, indicating that average time per move was not statistically different regardless of which problem was solved first. The main effect of solution $F(1,72) = 40.68$, $p < .001$, $MSE = 21{,}099{,}533.55$ was significant and the main effect of group $F(3,72) = 2.52$, $p = .064$, $MSE = 85{,}590{,}859.42$, approached significance. The Order x Group interaction was not significant $F < 1$, the three-way interaction $F(3,72) = 1.73$, $p > .05$, $MSE = 21{,}099{,}533.55$ and the Solution x Order interaction were not significant $F(3,72) = 2.94$, $p = .091$, $MSE = 21{,}099{,}533.55$, and the Solution x Group interaction was significant $F(3,72) = 3.06$, $p = .034$, $MSE = 21{,}099{,}533.55$. One potential explanation for the group and trend not yielding significant results could have been attributed to a lack of power, therefore potential explanations and follow-up analysis were discussed.

Independent samples *t*-tests conducted on the first solution revealed that those in the no-cost group ($M = 14{,}256.50$ milliseconds (ms)) took significantly less time to make moves when compared to the experience ($M = 21{,}665.85$ ms) and punishment groups ($M = 20{,}078.65$ ms) with $t(38) = 2.99$, $p = .005$ and $t(38) = 2.13$, $p < .05$, respectively. Independent samples *t*-test also revealed that the difference between the threat group ($M = 16{,}344.68$ ms) and the experience group, $t(38) = 1.93$, $p = .062$, approached significance. No additional differences were significant on the first or second solution. Figure 2-3 displays average move times for each group and solution.

The observed differences in move times may have indicated increased planning and/or attention in the experience and punishment groups, which could have potentially supported the CIF framework because increased checking and/or evaluation would have likely taken more time. Participants in the threat condition may have made moves quicker than those in the experience condition, although not statistically, which may have been due to a lack of power. In addition, the average times for the threat group did not differ significantly from those in the punishment group. These findings may have indicated that experiencing the penalty prior to beginning the problem had a greater effect of increasing attention and resulted in either more checking or greater evaluation of future moves. The lack of a significant difference between the no-cost and threat groups may have indicated that because the threat group did not experience the penalty the amount of increased attention generated from the threat may have been less than that generated in the experience condition.

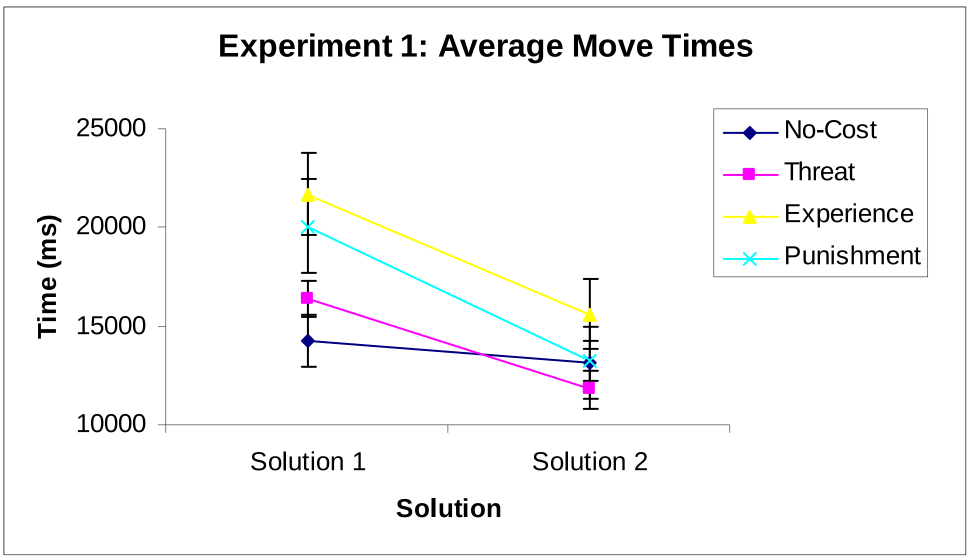

Figure 2-3. Average moves times in milliseconds by group for the first and second solution. Error bars represent standard errors.

An additional explanation for the greater move times in the punishment group could be attributed to the penalties, which occurred during the problem solving episode. After completing the penalty phase the participant may have had an adjustment period of re-engaging in the problem once he/she was returned to the last legal move. This re-engaging may have included recreating their previous move so that that specific illegal move could be avoided in the future or it may have involved recalculating all moves from that state. However, the greater times in the punishment group were more likely due to an increase in attention because similar increased move times were found in the experience group where an adjustment period explanation was not possible because the participant was not interrupted during the problem solution. In addition, the punishment group did not statistically differ from the threat group. The threat group may have also realized some increased attention, though not as much as the experience group. The lack of significant differences on the second solution could have been attributed to a floor effect where the participants were moving at a fairly efficient pace as a result of experience in completing the first solution.

Another explanation for the time differences could have been attributed to an error in the design of the computer program that was used. Because the timer was not initially set up properly to automatically "pause" during the penalty phase the generalized time adjustment (55 s for the initial penalty phase and 45 s for each subsequent penalty phase) may have led to significant variance compared to the actual times the participant took for penalties. The participant may have taken more or less than 10 s to learn the penalty rules in the experience and punishment conditions during the first penalty phase. The participant may have also taken additional time during each subsequent penalty in the punishment group before beginning the word rating task (words were not generated during this task until the participant clicked a button, a 45 s allotment for the penalty phase was based upon the assumption that the participant immediately clicked the word generation button as soon as the penalty screen appeared).

Paired samples *t*-tests for each group comparing average times between Solution 1 and Solution 2 were also conducted. This analysis revealed no difference between solutions for the no-cost group, t (19) = 1.08, $p > .05$, but significant differences for all other groups. In the threat group the participants were significantly faster on Solution 2 (M = 11,777.38 ms) than on the first solution (M = 16,344.68 ms), $t(19) = 2.95$, $p < .01$. In the experience group participants were quicker on Solution 2 (M = 15,614.67 ms) compared to Solution 1 (M = 21,665.86 ms), $t(19)$ = 4.13, p = .001. Finally, in the punishment group the second solutions (M = 13,257.43 ms) were also faster than the first solutions (M = 20,078.65 ms), $t(19)$ = 3.74, p = .001.

These results indicated that those participants that were penalized or were concerned with future penalty took longer to make moves on the first solution compared to the second, which would support the predictions of CIF that manipulations in Stage 1 would result in increased attention in Stage 2. Participants in the no-cost group, however, performed with similar speed on both problems, most likely because there was no penalty or concern of future penalties, consistent with CIF's prediction that a lack of consequence in Stage 1 would result in little or no increase in attention in Stage 2. Whether these results indicate increased planning, cautiousness, or some other mechanism was not something that could have been concluded by these results alone and required additional research. However, these results were promising and indicated that the threat of punishment, whether experienced or not, was sufficient to produce results similar to those observed when a penalty was administered after ever illegal move. These results lent support to the CIF framework that penalty and the threat of a penalty resulted in similar move times, potentially indicating an increase in attention.

Illegal move data. The illegal move data revealed no main effect of solution, $F < 1$, indicating that the participant made a similar number of illegal moves on the first and second problem. There was no main effect of order, $F < 1$, indicating that the participant made a similar number of illegal moves on each problem regardless of which problem was presented first. There was, however, a main effect of group $F(3,72) = 3.36$, $p < .05$, $MSE = 19.96$, indicating that participants in the four groups did not make the same number of illegal moves. Independent samples t-tests for the first solution revealed that the no-cost group made significantly more illegal moves on the first solution than the threat $t(38) = 2.37$, $p < .05$, experience $t(38) = 2.31$, $p < .05$, and punishment groups $t(38) = 3.15$, $p < .005$, which did not significantly differ from each other. Independent samples *t*-tests conducted on the second solution revealed that the groups did not differ from each other. Results are displayed below in Table 2-2 for illegal moves.

Although none of the two-way and three-way interactions were statistically significant it appeared that the trends indicated a potential 2-way Solution x Group interaction. On Solution 1, the no-cost group made significantly more illegal moves compared to the rest of the groups which did not differ. Participants in the punishment group ($M = 2.60$) made approximately one fewer illegal moves compared to the threat ($M = 3.40$) and experience ($M = 3.65$) groups, whose participants made a similar number of illegal moves. On the second solution, the number of illegal moves made by participants did not differ significantly between the four groups. However, participants in the threat ($M = 2.70$) and punishment ($M = 2.65$) groups made approximately two fewer illegal moves compared to participants in the no-cost ($M = 4.20$) and experience ($M = 4.65$) groups. An obvious explanation for these differences, which were not statistically significant, was that the experiment lacked the power to detect the effects with 20 participants in each of the four groups. Because these effects were potentially not detected due to a lack of power it did not seem unreasonable to expand upon these results.

Table 2-2. Illegal moves made on the first and second solutions were displayed by group and solution and included the means and standard deviations (SD). Superscript ^{3} is significant at $p < .005$, and all others significant at $p < .05$.

	Group	Mean	SD
Solution 1	No-Cost	$6.85^{1,2,3}$	5.82
	Threat	3.40^{1}	2.95
	Experience	3.65^{2}	2.16
	Punishment	2.60^{3}	1.60
Solution 2	No-Cost	4.20	4.41
	Threat	2.70	2.25
	Experience	4.65	7.65
	Punishment	2.65	2.08

A potential explanation for the trend that the punishment group made fewer illegal moves compared to the threat and experience groups on the first solution was that as these two groups made illegal moves and were not penalized the threat of penalty could have "worn off" over the duration of the solution. The participant could have started off avoiding illegal moves, but as there was no immediate penalty illegal moves could have increased as time increased. A second explanation for the trend could have been that the threat in these two groups was not as impactful as the immediate penalty in the punishment group and thus illegal moves rates were higher over the total duration of the problem. Trends on the second solution attempt were less clear with fewer potential explanations. The threat and experience groups were very similar except that those in the experience group completed the penalty phase once before beginning. Despite the overwhelming similarities between the two groups, including the similar number of illegal moves made on the first solution, participants in the experience group made approximately two additional illegal moves on the second solution. Due to the fact that the only difference between the groups was the completion of the penalty prior to beginning in the experience group, this was likely responsible for the trend. However, it remained unclear why completing the penalty once before beginning the problem on the first solution attempt would result in approximately two additional illegal moves on the second solution. Trends on the first and second solution were not significant and additional research would be required to determine if these trends would be significant with additional power and the cause of such results.

Paired samples *t*-tests revealed no differences between the first and second solution for the threat, experience, and punishment groups, all t's < 1. The paired samples *t*-test for the no-cost group was also not significant, but showed a trend towards a decrease in illegal moves on the second problem, $t(19) = 1.82$, $p = .084$. As was previously stated a lack of power may have potentially been responsible for some trends not being statistically significant. On the second solution (M = 4.20) participants in the no-cost group made more than two fewer illegal moves on average when compared to first solution (M = 6.85) performance. This appeared to be a large reduction in illegal moves, relatively speaking, but the trend was not statistically significant. This trend could have potentially been due to some type of general problem-solving learning on the first solution that allowed for transfer and illegal move reductions on the second solution. The illegal move changes from the first to second solution for the other three groups was also not significant, but the changes were smaller compared the no-cost group and seemed to be more in line with the statistical results. Because the *t*-tests between the first and second solutions for the other three groups were not significant this lent support to the idea that illegal move reductions on the second solution were sustained even after the penalty or threat of penalty had been removed.

The results in this section largely replicated and supported the results found in Knowles & Delaney (2005). The main effect of both solution and order were not significant and the no-cost group made significantly more illegal moves on the first solution compared to the punishment group. The punishment group, along with the threat group, did not show increases of illegal moves from the first solution to the second, potentially indicating sustained benefits through continued reductions in illegal moves. The experience group did not statistically show a change in illegal moves from the first to second solution, but the average did increase by one illegal move, 3.65 to 4.65, potentially indicating a lack of power to detect the change. In addition, participants in the threat and experience groups made significantly fewer illegal moves compared to the no-cost group and they did not significantly differ from the punishment group. These results provided support for CIF's assumptions: less intrusive methods of increasing attention and inducing caution were possible and resulted in similar illegal move reductions compared to the punishment group.

CHAPTER 3
EXPERIMENT 2

The purpose of the second experiment was to determine whether or not a punishment or threat of punishment was necessary to induce cautiousness. The two critical groups in this experiment received no negative consequences for illegal moves. Instead one group was instructed to “do their best” and to concentrate on avoiding illegal moves and the other group was rewarded for avoiding illegal moves and performing efficiently (performing efficiently was explained later in this chapter). Similar to Experiment 1, both of these groups were compared to a group that was punished for illegal moves and a group that received no punishment. In this experiment there were no negative consequences or threat of negative consequences for the illegal moves in the two critical conditions. Therefore, the results should have provided insight into whether or not negative consequences were necessary to induce caution.

As in Experiment 1, Stage 1 was manipulated, but this time the consequences were a reward or motivating instruction (as opposed to the threats used in Experiment 1). This manipulation would have then affected Stage 2, where motivation would have been switched “on” and attention would have been increased according to CIF. This increase in attention would have ultimately resulted in a reduction in illegal moves. The author predicted that there would have been fewer illegal moves in the motivation and reward groups compared with the no-cost condition that was not punished and did not receive additional instructions or reward. It was also predicted that the reward condition would likely reduce illegal moves to a greater degree than the motivating instruction. If a reward or motivating instruction was able to induce cautiousness then this would have supported the claim of the framework; manipulations in Stage 1 influenced people to increase their attention and this was what caused a reduction in illegal moves, not punishment or fear of punishment, but an increase in attention to the problem.

Methods

Participants

The process for recruiting participants was the same as the first experiment. In Experiment 2, 108 participants took part in the study; each was randomly assigned to one of the four groups. A total of 28 participants were unable to solve both problems correctly within the time limit and their data was therefore dropped from the analysis. Participants were that were unable to solve the problems were replaced until there were 20 participants in each of the four groups.

Problems

The same problems were used as in Experiment 1 -- hobbits & orcs and titration.

Design

Four groups were used in this experiment: no-cost, punishment, motivation, and a reward group. As in Experiment 1, the groups differed in the consequences of making illegal moves on the first of two problems. All groups received identical instructions when they solved the second problem – specifically, they were informed that there was no consequence or reward for avoiding illegal moves. The purpose of fixing the instructions for all four groups on the second problem was to examine transfer effects. The manipulations on the first problem were as follows:

The *no-cost* group was identical to that used in Experiment 1, and received no special instructions, rewards, or penalties for making illegal moves.

The *punishment* group differed only slightly from the Experiment 1 punishment group. The punishment (rating the words for pleasantness) occurred on a fixed interval schedule rather than immediately after every illegal move. The participant had to complete the penalty every 30 s if a penalty was made in that time frame. When an illegal move was made, the participant saw a message box stating that an illegal move was made and then they were allowed to continue working on the problem from the last legal state. Once the 30 s time interval was complete the screen then turned gray and the participant was instructed that he/she recently committed an illegal move and he/she would be instructed to complete the penalty. If an illegal move was not made in the 30 s interval then the participant received no information and continued working on the problem until the end of an interval in which an illegal move was committed, the problem was solved, or time expired. Administering the penalty at a fixed interval instead of immediately after every illegal move was done to maximize the similarity between the punishment and reward conditions.

In the *motivation* condition, prior to starting the first problem, the participant was instructed that the main focus of the experiment was to assess his/her ability to avoid illegal moves. The participant was informed to “do their best” to avoid illegal moves in an attempt to solve the problem with the fewest number of illegal moves possible. No reward, penalty, or additional instructions were given to the participant in this group.

Finally, in the *reward* condition, prior to starting the first problem the participant was asked for his/her preference of Skittles or M&M candies. The participant was then instructed that he/she would be rewarded with the candy of choice for avoiding illegal moves and solving the problem efficiently. The participant was told that the reward would be based primarily on the avoidance of illegal moves, but to maximize the reward he/she should try to complete the problem in a timely manner and in the minimum number of moves possible. A dish was next to the participant and the experimenter deposited one piece of candy into the dish every 30 s if no illegal moves were made in that time frame and the participant reached a new legal state in that time frame that had not been previously visited during the problem solving episode. If in the 30 s time frame the participant reached a new legal state and then back-tracked to a previously visited state the participant was still rewarded with candy as long as an illegal move was not made in that interval. In addition, the participant received an additional candy for completion of the problem and an additional piece of candy for every 30 s interval remaining before six minutes (e.g. three extra candies for completing the problem before 4:30 because there were three full 30 s intervals remaining before six minutes was reached). Six minutes was chosen based on the average time to complete the problem obtained in Knowles and Delaney (2005).

Procedure

The procedures followed that of the first experiment with the exception that those in the motivation and reward condition received additional instructions before beginning the first problem. After completing the first problem, participants in the motivation and reward conditions were asked if they believed they were going to receive punishment for making illegal moves. As in Experiment 1, participants completed the second problem under no-cost instructions.

Results and Discussion

Move data. A 4 x 2 x 2 mixed factorial ANOVA was conducted: Group (No-Cost vs. Punishment vs. Motivation vs. Reward) x Order (Hobbits first vs. Titration first) x Solution (First vs. Second). Group and order were between-subjects factors and solution was a within-subjects factor. As in the first experiment, illegal moves committed on the first and second problem were of primary importance, but number of legal moves and time per move were also analyzed.

Based on previous findings obtained by Knowles and Delaney (2005) it was anticipated that there would be no main effect of order, no main effect of solution, and no interaction for any of the factors. This would indicate that the isomorphs did not differ in difficulty, the order of presentation did not influence the results, and the results for the first solution did not differ significantly from the second solution. It was also anticipated that legal moves and average time per move would not differ significantly between the four groups. However, it was predicted that a main effect of group would be obtained for illegal moves.

Legal move data. The results for legal moves revealed no main effect for solution $F < 1$ or order $F(1,72) = 1.20$, $p > .05$, $MSE = 286.19$. Although the result of the analysis on group revealed no main effect $F(3,72) = 2.42$, $p = .073$, $MSE = 286.19$, it did approach significance. The trend for this analysis was for participants in the reward group to make the most legal moves ($M = 32.03$) and for participants in the no-cost ($M = 24.63$), motivation ($M = 23.35$), and punishment ($M = 23.40$) groups to make similar numbers of legal moves. This trend, though not significant, could have potentially been due to participants adapting a strategy of trying to maximize rewards by navigating through legal states while avoiding illegal moves, thus increasing legal moves. Previous results have indicated that legal moves have remained fairly stable during experimental manipulations so this trend was somewhat surprising.

Move time data. The procedures for calculating average times per move were the same as those used in Experiment 1. The punishment group was adjusted 55 s for the first penalty and 45 s for each subsequent penalty. Penalties occurred at a fixed interval so the number of penalties may have been fewer than the number of illegal moves. If more than one illegal move occurred in a given interval, time was adjusted per penalty phase. The results for move time data revealed no main effect for solution or order, both *F's* < 1. However, the main effect of group was significant, $F(3,72) = 2.77$, $p < .05$, $MSE = 29{,}883{,}813.92$, the Solution x Order interaction was significant, $F(1,72) = 4.70$, $p < .05$, $MSE = 12{,}203{,}979.10$, and the Order x Group interaction approached significance, $F(3,72) = 2.66$, $p = .054$, $MSE = 29{,}883{,}813.92$. The means and standard deviations for each group by solution were displayed in Table 3-1.

Independent samples *t*-tests revealed that the punishment group took longer to solve the first problem compared to the reward and motivation groups, as displayed in Table 3-1. These results could have been attributed to a more “carefree” mentality in the reward and motivation condition due to the motivating instructions or potential reward of candy and potentially increased planning in the punishment group as participants attempted to avoid illegal moves. In addition, the penalty in the punishment group occurred at a fixed interval after an illegal move was made in that interval. So, an alternative explanation could have been that participants were waiting or preparing as they knew the penalty phase could be initiated at any time and this resulted in increased move times. A final alternative proposed in the first experiment was that after completing a penalty phase participants may have required a re-adjustment period when returning to the problem and this would have increased move times.

Table 3-1. Means and standard deviations (SD) by group and solution for average times per move were presented in milliseconds. Superscripts [1, 2, 3] were significant at $p < .05$.

	Group	Mean	SD
Solution 1	No-Cost	14124.72	5055.92
	Motivation	11593.93[1]	3812.97
	Reward	12083.44 [2]	3480.25
	Punishment	14938.61 [1,2]	4186.00
Solution 2	No-Cost	13086.21	4252.88
	Motivation	12892.55	5652.86
	Reward	10961.40 [3]	3026.99
	Punishment	14578.37 [3]	6916.07

On the second solution, the reward group remained significantly quicker than the punishment group, but the difference between motivation and punishment was no longer significant. It was possible that participants in the punishment group were still being cautious and taking additional time planning or preparing for a penalty that could have appeared at any moment. Participants in the reward condition may have taken the problem less seriously and been less cautious due to the fact that they were administered candy on the first solution. The distribution of candy on the first solution may have made the problem seem like more of a game than an experiment resulting in less attention, planning, and/or caution.

The order by group interaction approached significance and post hoc analysis were conducted as the study design may have lacked the power to detect the effect. An independent samples *t*-test analysis revealed that participants in the no-cost group may have had longer average move times when solving the titration problem first (M = 15,325.77) compared to participants that solved the hobbits and orcs problem first followed by the titration problem (M = 11,885.17), $t(18) = 2.07$, $p = .053$ as this analysis approached significance. Results from the independent samples t-tests for the motivation $t(18) = 1.18$, $p > .05$, reward $t(18) = 1.67$, $p > .05$, and punishment groups $t(18) = 1.15$, $p > .05$, revealed non-significant results. One potential explanation for this trend could was the idea that those in the no-cost group had no consequences for performance and therefore no reason to devote additional attention towards the problem. This lack of additional resources toward the problem may have resulted in more difficulty encoding the titration problem when it was presented first, as it may have appeared more abstract and provided less information to the participants (participants were unable to see the contents of the destination beakers in the titration problem whereas participants could always see both banks of the river in the hobbits and orcs problem). Average move times for the different groups by order were presented in Figure 3-1.

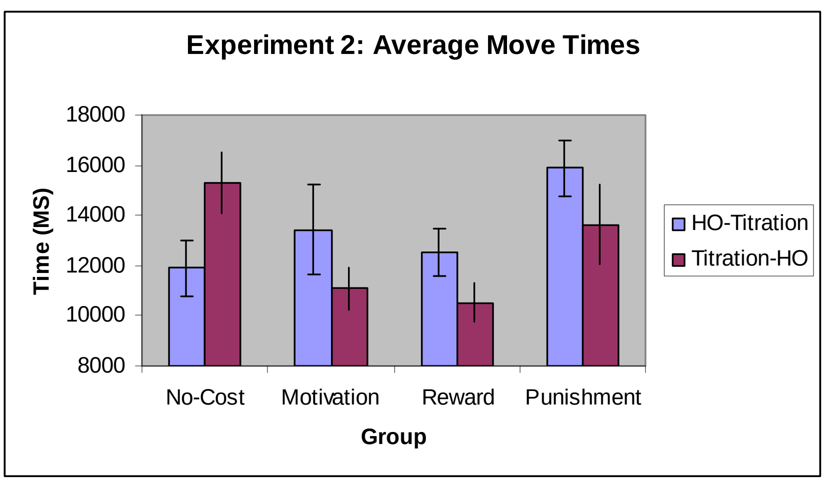

Figure 3-1. Average time per move for each of the four groups was displayed for participants that solved the hobbits and orcs (HO) problem first and those that solved the titration problem first. The error bars represented standard error.

Illegal move data. The analysis of the illegal moves made revealed no main effect of solution, $F(1,72) = 2.77$, $p = .099$, $MSE = 10.99$, no main effect of order, $F < 1$, and no main effect of group, $F(3,72) = 1.86$, $p = .321$, $MSE = 19.02$. In addition, the Solution x Order and Group x Order interactions were not significant, both *F*s < 1. Neither the Group x Solution interaction, $F(3,72) = 2.23$, $p = .092$, $MSE = 10.99$, nor the three-way interaction were significant, $F(3,72) = 1.57$, $p = .205$, $MSE = 10.99$. These results were surprising given the previous findings from Knowles and Delaney (2005) and Experiment 1, in which the punishment group made significantly fewer illegal moves compared to the no-cost group on the first solution. Table 3-2 provides the means and standard deviations for illegal moves by group and solution.

One potential explanation for the absence of significant results could have been due to a lack of power. For this reason, independent samples *t*-tests were conducted between the groups for illegal moves on the first solution attempt. The independent samples *t*-tests comparing no-cost and motivation, $t(38) = 1.98$, $p = .056$, no-cost and reward, $t < 1$, no-cost and punishment, $t(38) = 1.24$, $p = .222$, motivation and reward, $t(38) = 2.00$, $p = .053$, motivation and punishment, $t(38) = 1.01$, $p = .320$, and reward and punishment, $t(38) = 1.55$, $p = .129$, were not significant. The punishment group has typically made fewer illegal moves compared to the no-cost group, but this comparison was not significant. Also of interest were the two experimental groups compared to the no-cost group, which revealed a trend approaching significance comparing the motivation and no-cost groups. Although not significantly, the motivation group made the fewest number of illegal moves ($M = 3.40$) out of all four groups. It may have been possible that with additional power both the punishment and motivation groups would have made significantly fewer illegal moves compared to the no-cost group.

The most unexpected finding in this experiment was the lack of a significant difference in illegal moves between the no-cost and punishment group. One potential explanation could have been that since the penalty was set at a fixed interval, the number of penalties could have been (and often were) fewer than the number of illegal moves committed. For example, if the participant committed an illegal move at the beginning, the middle and the end of a 30 s interval then three illegal moves were made, but only one penalty was administered at the end of the 30 s time interval. Being penalized less often compared to previous experiments could have reduced the effect, resulting in non-significant differences between the groups for illegal moves.

Table 3-2. Means and standard deviations (SD) of illegal moves by group and solution for illegal moves were presented.

	Group	Mean	SD
Solution 1	No-Cost	5.40	3.42
	Motivation	3.40	2.96
	Reward	6.85	7.13
	Punishment	4.25	2.34
Solution 2	No-Cost	3.75	2.55
	Motivation	3.95	2.63
	Reward	4.15	3.10
	Punishment	4.55	4.43

Also of interest and importance was the lack of a significant difference between the no-cost group and the two experimental conditions. One explanation for the lack of a reduction in illegal moves in the reward condition could have been attributed to the reward itself. It was possible that candy was not very rewarding to the participants and significant improvements may have been obtained with a more valuable reward, like money. An additional possibility was that the reward rules awarded additional or bonus candy at the completion of the problem for efficient performance and this may have had little or no effect on problem solving behavior. Awarding candy after the problem had been completed may not have influenced participants to avoid illegal moves during the problem. In addition, the lack of a reduction of illegal moves in the motivation condition could have indicated that the instructions were not effective in increasing attention.

Given that this experiment was run last out of the three experiments and was run at the end of the semester this may have resulted in less motivated students, which may have affected the results. Students who had waited until the end of the semester to volunteer and complete their experimental credits were likely less motivated to complete the credits and may have been more resistant to the experimental manipulations. Another explanation for the results could have been that an adverse stimulus or threat of an adverse stimulus was required to increase attention and reduce illegal moves. Thus, participants in the motivation and reward conditions would not have been influenced to increase attention and reduce illegal moves. Unfortunately, the lack of typical effects in this experiment could not be clearly understood from the collected data and further research is required to answer these questions.

CHAPTER 4
EXPERIMENT 3

The previous two experiments concentrated on finding ways to aid participants in becoming more attentive, through threat or reward, so that they could perform more efficiently on the problem. Once a participant had become motivated and decided to allot additional attention to the problem, what did this increased attention do to reduce illegal moves? In previous research Knowles and Delaney (2005) proposed that punishment induced caution and that caution was checking a move against the rules for legality so that illegal moves could be rejected and avoided. Experiment 3 instructed participants to engage in this checking behavior so that illegal moves could be avoided and so that the result of such behavior could be assessed.

In the critical conditions of Experiment 3, participants were asked to check the legality of each move after selecting it, but before executing it. One group was asked to perform this check while thinking aloud so that it could be monitored by the experimenter and the other group was asked to perform the check silently. Participants received training in thinking aloud and were asked to verbalize any thoughts they had while working on the problem. The training and instructions were similar to those proposed by Ericsson and Simon (1993). The silent check group was used for comparison to the aloud group to ensure that the aloud task did not affect performance. Based on previous research, the illegal move rates of the silent and aloud check groups should not have differed; both should have shown significant reductions in illegal moves and all other data between the two groups, except for time per move, should have been similar (Ericsson and Simon, 1993; Knowles and Delaney, 2005).

Instructing a participant to check a move for legality would have likely increased the frequency of this checking behavior. This would have ultimately influenced participants to reject more illegal moves compared to those not instructed to check. Asking participants to check moves for legality would have affected Stage 2 of CIF because participants would have become more attentive with the requirement of checking each move for legality. Platt and Griggs (1993) found that when participants were asked to explain their choices in Wason's 4-card selection task they obtained the highest success rates every reported for this difficult task. These results appeared to support the idea that increasing engagement in a task, whether through requiring explanations for behavior or through possibly a check of legality, can yield significant improvements in performance. Instructing participants to check moves would also have a direct effect on Stage 3 of the framework where checking behavior was directly manipulated.

Instructing participants to check the legality of each move before it was made allowed CIF to make clear predictions. CIF predicted that the instruction to check would have increased attention and decreased the number of illegal moves made compared to the no-cost group. The prediction was that illegal move reductions would have achieved rates similar to the punishment group in the first experiment and previous findings of Knowles and Delaney (2005).

In addition, a measure of working memory using the operation span (OSPAN) task was also administered at the end of the experiment (Kane, Bleckley, Conway, & Engle, 2001; Turner & Engle, 1989). The reason for obtaining working memory scores was to determine the extent to which resources play a role in human problem solving. If the resource limitation hypothesis (Jeffries et al., 1977) was correct then working memory should significantly correlate with the number of illegal moves a participant makes. The OSPAN task assessed working memory by presenting equations and words on a computer screen. The participant had to verify the equations while maintaining the words in memory. To assess the participant's working memory span the words had to be recalled in the order that they were presented. This task was described in more detail in the procedure section.

The purpose of measuring OSPAN was to examine the role played by working memory resources in problem solving, which has been proposed as a cause for the selection of illegal moves (Jeffries et al., 1977). Those with low working memory spans may have had difficulty calculating future states or remembering to check each move for legality contributing to the difficulty of the problem. Previous research on these types of problems did not reveal a significant relationship between illegal or legal moves made and working memory (Knowles and Delaney, 2005). However, the evidence here may prove otherwise since participants were instructed to check aloud. This manipulation allowed the experimenter to verify that the participant actually engaged in this behavior and enabled the experimenter to assess the rate at which they checked and the accuracy of this checking. The role of working memory in problem solving was outside the scope of CIF; therefore the framework was unable to make a prediction as to the relationship between problem solving and working memory. The author believed that such factors contributed to problem solving and predicted a negative correlation between illegal moves and working memory and a negative correlation between legal moves and working memory.

Methods

Participants

The process for recruiting participants was the same as the first two experiments. In Experiment 3, 103 participants took part in the study; each was randomly assigned to one of the four conditions. Of these, 23 participants were unable to solve both problems correctly within the time limit and their data was therefore dropped from the analysis. Participants that were unable to solve the problems were replaced until there were 20 participants in each of the four groups.

Problems

The problems used in Experiment 3 were the same hobbits & orcs and titration problems as those used in the first two experiments.

Design

In this experiment there were four groups: no-cost group, a punishment group, a check group, and an aloud group. As in Experiments 1 and 2, the groups differed as to the instructions they received on the first problem and the instructions on the second problem did not differ between the groups so that transfer effects could be examined. All groups solved the second problem silently without punishment or instructions to check each move.

The punishment and no-cost groups were identical to those in Experiment 1. The *no-cost* group received no special instructions, rewards, or penalties for making or avoiding illegal moves. The *punishment* group penalized participants immediately after every illegal move by having them rate the pleasantness of words for 45 s just as participants did in the first experiment.

The *check* condition instructed the participant that after selecting a move he/she would check the move for legality before executing that move. A button was located on the display so that after checking the move but before executing it, the participant would click the button to indicate that they had checked the move and that he/she believed it to be a legal move. This button was primarily on the display so the experimenter could verify that those in the silent condition were performing the check. In the aloud condition the button remained on the display so that the similarity between the two conditions was maximized. Participants in the aloud condition were also instructed to click on the check button before making a move to indicate that the move had been checked. In addition, the computer program tracked when and how often a participant forgot to click the button to indicate that they had checked the move for legality.

A participant in the *aloud* condition received instructions and training on thinking aloud and was then asked to verbalize his/her thoughts as they worked on the problem (instructions for thinking aloud were provided in the Appendix). These procedures were similar to the instructions proposed by Ericsson & Simon (1993). If a participant was silent for more than 3 s or speaking too softly he/she was reminded to keep talking. The experimenter attended to the participant's verbal protocols in an attempt to ensure that the participant checked each move for legality before executing that move. As in the check condition the participant was asked to click on the "Check" button before finishing his/her move to indicate that he/she had checked the move and believed it to be a legal move. If the participant did not check a move for legality he/she was reminded to continue doing so and the computer tracked each checked and non-checked move. The participant was asked to speak into a microphone while solving the first problem and verbal protocols were recorded on a mini-disc recorder.

Procedure

The procedures followed those of Experiment 1, except that participants in the aloud condition received instructions and training at the beginning of the tutorial phase, to think aloud as they solved the problem. The participant was given instructions for thinking aloud similar to those suggested by Ericsson and Simon (1993) and was asked to verbalize his/her thoughts by saying whatever came into his/her head. The participant was then instructed to think aloud as he/she imagined him/herself walking through a house that he/she was very familiar with and to count the number of windows in the house. The experimenter stopped this task when the participant was able to provide descriptors of the interior of the house, which demonstrated understanding of thinking aloud. If the participant was too vague or did not understand the task then the experimenter provided an example of describing the interior of a house while counting the windows and then asked the participant to complete the task. This task was very brief and typically did not take more than 30 s; every participant in the think aloud group was able to complete this task. During the tutorial phase, participants in the aloud group were asked to work on the practice problem while thinking aloud and to check each move for legality. The other three groups completed the tutorial phase in a fashion similar to the participants in Experiment 1. As in the other experiments, all participants solved the second problem silently and without checking. After completion of both problems, participants in all groups completed the operation span (OSPAN) task to assess working memory capacity (Kane, Bleckley, Conway, & Engle, 2001; Turner & Engle, 1989).

The OSPAN task was a task used to assess a person's ability to maintain information in memory while completing another task, which evaluated a person's working memory span. The task presented a simple mathematical equation followed by a word (e.g., (9 / 3) – 1 = 2 CONE) in a Power Point presentation on a computer screen. The participant was to read the equation aloud, verify whether or not the equation was correct (some equations were true and some were false), and then say aloud the word that followed the equation. Once this was completed the experimenter immediately presented the next slide on the computer screen with a new equation and word. Once the new slide appeared the participant was to begin reading the equation aloud immediately, verify it and then read the word. Anywhere between two and five equations/words were presented and then the participant saw a colored screen with "???" at the completion of a *series* (all the words presented since the last colored screen). This was the participant's cue to recall all the words he/she had seen in that series on a piece of paper in the order in which they were presented. There were a total of 12 series (three series each with two words, three words, four words, and five words) and they were randomly presented so that the participant did not know how many words would be in that series until that series was over. The participant received one point for every correct word, although the participant only received points if all the words in the series were recalled correctly in the correct order (e.g., no points given if all four words recalled, but the order of Words 1 and 2 were switched; four points given if all words recalled correctly in the correct order).

Results and Discussion

Move data. For this experiment a 4 x 2 x 2 mixed factorial ANOVA was conducted: Group (No-Cost vs. Punishment vs. Check vs. Aloud) x Order (Hobbits first vs. Titration first) x Solution (First vs. Second). Group and order were between-subjects factors and solution was a within-subjects factor. As in the first two experiments, illegal moves committed on the first and second problem were of primary importance, but number of legal moves and time per move were also analyzed.

As in the first two experiments it was anticipated that there would be no main effect of order, no main effect of solution, and no interaction for any of the factors indicating that the isomorphs did not differ in difficulty, the order of presentation did not influence the results, and the results for the first solution did not differ significantly from the second solution. It was also anticipated that legal moves and average time per move would not differ significantly between the four groups. Those participants in the check and aloud conditions could have had slightly higher times as a result of the instruction to check and the process of verbalizing thoughts in the aloud condition, but this should not have affected performance otherwise. However, it was anticipated that a main effect of group would be obtained for illegal moves.

Legal move data. An analysis of legal moves revealed no main effect of solution, $F < 1$, no main effect of order, $F(1,72) = 1.31$, $p > .05$, $MSE = 228.21$, and no main effect of group, $F(3,72) = 1.07$, $p > .05$, $MSE = 228.21$. These results indicated that, regardless of which problem was solved first, the number of legal moves did not differ between the first and second solution or between the four groups. These results replicated the findings of the first two experiments and support the claim that the manipulations have little or no effect on the legal moves within the problems.

Move time data. Procedures for calculating average time per move were the same as those used in the first two experiments, and the adjustments made to the punishment group for illegal move penalties were the same as those imposed on data from Experiment 1. Analysis revealed no main effect of order, $F(1,72) = 2.60$, $p > .05$, $MSE = 50{,}735{,}853.35$. However, a main effect of solution was found, $F(1,72) = 42.10$, $p < .001$, $MSE = 37{,}739{,}536.72$, which indicated that participants took less time making moves on the second solution ($M = 12{,}645.75$ ms) compared to the first solution ($M = 18{,}948.23$). A main effect of group, $F(3,72) = 3.97$, $p = .011$, $MSE = 50{,}735{,}853.35$ was present, indicating that the groups did not make moves at the same rate. The results also revealed a two-way Solution x Group interaction, $F(3,72) = 5.64$, $p = .002$, = 37,739,536.72. As indicated in Figure 4-1 some of the times decreased from the first solution to the second and some did not. Paired samples *t*-tests revealed no change for the no-cost group from the first solution to the second, $t(19) = 1.65$, $p > .05$. The difference for the check group, $t(19) = 2.27$, $p < .05$, the aloud group, $t(19) = 9.20$, $p < .001$, and the punishment group, $t(19) = 2.19$, $p < .05$, showed significant decreases in time from the first solution to the second. The means for the groups by solution attempt were displayed in Figure 4-1.

Independent samples *t*-tests were also conducted on the groups for the first and second solutions. There were no significant differences between the groups for the second solution attempt. On the first solution attempt the no-cost group took less time on moves compared to the check, $t(38) = 2.84$, $p < .01$ and aloud groups, $t(38) = 5.86$, $p < .001$, and approached significance for the punishment group, $t(38) = 1.70$, $p = .098$. The check group did not significantly differ from the aloud, $t(38) = 1.60$, $p > .05$, or the punishment groups, $t < 1$. The punishment group took significantly less time making moves on the first solution when compared to the aloud group, $t(38) = 2.50$, $p < .05$. Average move times for the groups were displayed in Figure 4-1 by solution attempt.

The check, aloud, and punishment groups took significantly less time on the second solution compared to the first. The results may have been due to the increased activities such as checking and/or thinking aloud in the experimental conditions, but may have also been due to increased attention as proposed in Experiment 1. It was also not surprising that the check and aloud groups took significantly longer to make moves on the first solution when compared to the no-cost group because the two experimental groups had to take additional actions in checking moves for legality and/or speaking aloud. During the second solution when the manipulations were removed all groups had similar move times.

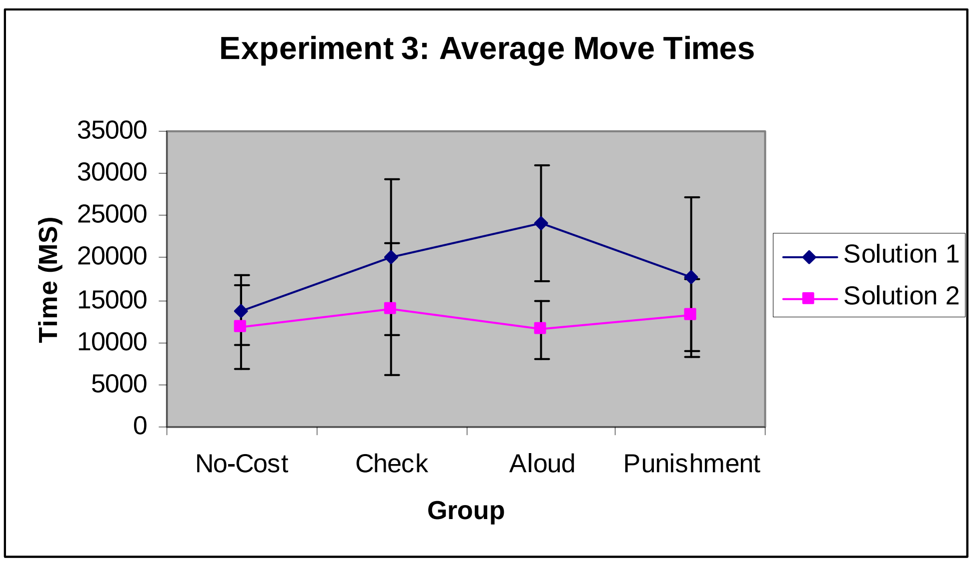

Figure 4-1. Average time per move for each of the four groups was displayed for the first and second solutions in milliseconds. The error bars represented standard error.

Illegal move data. An analysis of illegal moves revealed that there was neither a significant main effect of solution, $F < 1$, nor a significant main effect of order, $F < 1$. The main effect of group was significant, $F(3,72) = 3.12$, $p < .05$, $MSE = 18.24$, and was followed up with independent and paired samples *t*-tests. Independent samples *t*-tests revealed that the no-cost group made significantly more illegal moves when compared to the check, $t(38) = 2.59$, $p < .05$, aloud, $t(38) = 3.41$, $p < .005$, and punishment group, $t(38) = 3.03$, $p < .005$. The paired samples *t*-tests revealed no significant differences between solutions for the no-cost, check, and punishment groups. The analysis of the aloud group indicated that participants made more illegal moves on the second solution when compared to the first. Figure 4-2 displayed the group means for both the first and second solution. In addition, the Solution x Order and 3-way interactions were not significant, *F*s < 1. The Solution x Group interaction, $F(3,72) = 1.94$, $p =.132$, $MSE =$ 12.20, and Order x Group interaction, $F(3,72) = 1.27$, $p =.291$, $MSE = 18.24$, were also not significant. Illegal moves for each group for the first and second solutions were displayed in Figure 4-2.

The analysis of illegal moves replicated previous findings that when participants were penalized for illegal moves they made significantly fewer illegal moves compared to participants that were not penalized for making illegal moves. In addition, participants in the two experimental conditions made significantly fewer illegal moves compared to the no-cost control group. These findings indicated that when participants were instructed to be and were accountable for checking moves for legality this resulted in the reduction of illegal moves. These results directly addressed and supported Stage 3 of CIF where an increase in checking behavior resulted in a decrease in illegal moves. This reduction occurred even when participants were instructed to think aloud while solving the problem.

One very interesting result was that participants in the aloud group made significantly more illegal moves on the second solution ($M = 4.60$) than they did on the first ($M = 2.00$), $t(19)$

= 2.73, p = .013. Even though the check and punishment groups showed sustained benefits on the second solution with no significant increases in illegal moves, the aloud group did not and showed a release, making more illegal moves on the second solution than the first.

An increase of illegal moves from the first solution to the second for the aloud group was an unexpected result. Knowles and Delaney (2005) performed a similar experiment where participants were instructed to think aloud under no-cost instructions or to think aloud under punishment instructions. Neither group was instructed to perform a check of moves for legality. Both groups in Knowles and Delaney's experiments showed no significant effects of thinking aloud and showed typical no-cost and punishment results, but neither group solved a second problem to assess transfer effects. In addition, the check group in the current experiment showed sustained benefits with no significant difference between illegal moves on the first and second solution. Since transfer effects were not previously explored with think aloud instructions, these results could be due to the think aloud instructions or to some interaction between thinking aloud and checking.

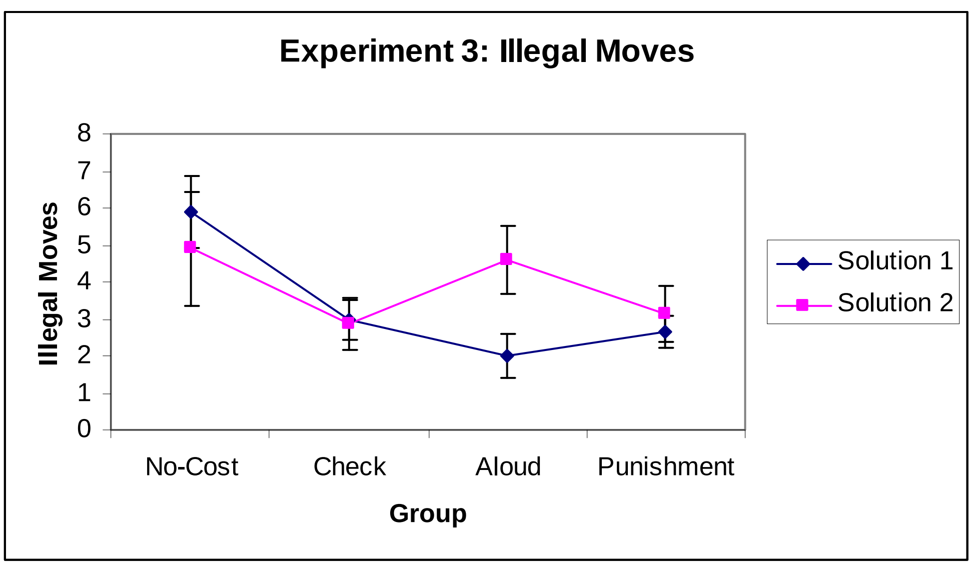

Figure 4-2. Illegal moves for each group were presented by condition with standard errors indicated as error bars.

Checking data. The checking data provided insight into the relationship between checking behavior and the reduction of illegal moves, a direct indication of how checking affected illegal moves. Participants in the check and aloud groups were instructed to click on a "check" button on the computer display before executing a move to indicate that they had checked that move for legality. The computer program tracked when a move was not checked prior to making that move. This information allowed for a physical indication of when a participant did not check a candidate move for legality. These values were compared with the number of illegal moves committed by each participant in the check and aloud groups. Illegal moves were not reliably correlated with checking in the check group, $r(19) = .244$, $p = .300$. In the aloud group, as the number of illegal moves decreased the frequency of missed checks decreased, $r(19) = .657$, $p = .002$.

The results from the correlation analysis revealed some unexpected results in that checking in the check group was not significantly related to illegal moves, but checking in the aloud group was related to illegal moves. One potential explanation for this finding was that those in the check group did not verbalize their thoughts and therefore were not held accountable to the same degree as those in the aloud group. A participant in the check group could have simply clicked the "check" button prior to each move without actually engaging in any checking behavior because there was no way for the experimenter to actually verify what the participant was doing "in his/her head." This explanation seemed unlikely because participants in the check group made a similar number of illegal moves compared to participants in the aloud group and fewer than those in the no-cost group. In contrast, a participant in the aloud group was reminded to continue verbalizing if he/she was silent for too long, which may have influenced the participant to be more compliant in checking moves. An additional explanation could have been that participants in the check group could have checked moves as often as those in the aloud group, but forgot to click the check button more often as a result of not verbalizing. The results did not support this explanation -- the number of times the check button was not clicked prior to making a move did not statistically differ between the check (M = 1.75) and aloud (M = 2.40) groups, t < 1. The check and aloud groups both made fewer illegal moves compared to the no-cost group and did not differ from each other, which may have indicated that participants in the check group were likely checking and rejecting moves to the same degree as the aloud group.

OSPAN data. After completion of the two problems, participants were asked to complete the OSPAN task to assess their working memory span. Out of the 80 total participants that solved both problems within the 20 min time limit in Experiment 3, 10 participants were unable to complete the OSPAN task in the time allowed. The other 70 completed the OSPAN task and their data were submitted to a correlation analysis. Nineteen participants in the check group completed the OSPAN task, but their scores were not significantly related to either the number of illegal moves made $r(18) = -.067$, $p = .785$ or the number of missed opportunities to check a move for legality $r(18) = -.252$, $p = .297$. Eighteen participants in the aloud group completed the OSPAN task, their scores were not significantly related to the number of illegal moves made $r(17) = -.275$, $p = .269$, but they were significantly related to the number of missed opportunities to check a move for legality $r(17) = -.479$, $p = .044$. The measure obtained by the computer program was the number of times a participant made a move without first clicking the check button, in other words, the number of missed opportunities to check for legality. This negative correlation indicated that as a participant's working memory span score increased the probability of that participant forgetting to check a move for legality decreased.

The correlation between missed opportunities to check a move for legality and memory span score in the aloud group appeared to support the resource limitation hypothesis. That is, those participants with more available memory were more able to remember to check moves for legality. However, a similar finding was not found in the check group and the correlations between memory span scores and the number of illegal moves made were not significant, which did not support a resource limitation hypothesis. One potential explanation could have been that with the instruction to think aloud and check moves for legality simultaneously in the aloud group this could have taxed enough resources that those with larger working memory spans were more able to perform the check consistently when compared to those with smaller working memory spans. This explanation did not seem to explain why similar benefits were not seen in correlations between working memory and illegal moves for the aloud group. Taken as a whole these results seemed to indicate that a resource limitation hypothesis could not be entirely ruled out, but that it likely played a smaller role than has been assumed in previous research.

CHAPTER 5
GENERAL DISCUSSION

The three experiments presented here addressed specific assumptions of the CIF framework for the reduction of illegal moves. Experiment 1 tested whether an intrusive penalty was necessary to reduce illegal moves or if a threat without penalty for each illegal move would be sufficient. Experiment 2 set out to determine if an aversive stimulus was necessary to reduce illegal moves or if a more positive consequence could help to improve problem solving efficiency. Experiment 3 instructed participants to engage in checking behaviors to directly assess the ability of these actions to reduce illegal moves. Experiments 1 and 3 seemed to support aspects of the framework, while Experiment 2 produced non-significant results. As noted previously, several of the trends seemed intriguing, but low statistical power to detect some of these effects was likely at fault. Therefore, some of the trends that were not significant were expanded upon further in this section.

Illegal Moves

The first experiment demonstrated that the threat of a penalty, whether it was experienced or not, was able to reduce illegal moves compared to a group that was not penalized or threatened. Participants in the no-cost group made the most illegal moves (M = 6.85) and the threat and experience groups made fewer illegal moves (Ms = 3.40 and 3.65, respectively), but similar numbers compared to each other. Because the threat and experience groups made similar numbers of illegal moves, verbal instructions alone without any corporeal consequences were apparently sufficient to reduce illegal moves. Participants in the punishment group made slightly fewer illegal moves (M = 2.60) compared to the two experimental groups on the first solution, although not significantly. One potential explanation for the trend of fewer illegal moves in the punishment group could be attributed to the idea that the punishment manipulation was more powerful and thus had the potential to reduce illegal moves to a greater degree compared to a threat alone. The lack of significance may have also been attributable to a floor effect, whereby significance may have been obtained in problems with a greater opportunity to make more illegal moves.

Surprisingly, the second experiment did not reveal any significant illegal move reductions on the first solution. The intent of the motivation and reward manipulations was to determine if positive consequences or instructions to avoid illegal moves would result in reductions. The motivation (M = 3.40) and punishment (M = 4.25) groups, although not significantly, made fewer illegal moves compared to the no-cost group (M = 5.40). It is possible that with increased power to detect the effects the motivation and punishment groups would have made significantly fewer illegal moves compared to the no-cost group. The trend indicated that participants in the motivation group made the fewest illegal moves. Why they potentially made fewer than the punishment group is not well understood. One possibility was that the extra motivating instructions may have seemed suspicious to participants and caused them to be even more cautious for fear of the "unknown." Alternatively, they may have taken the instructions as a hint that avoiding illegal moves was the secret to solving the problem.

Candy in the reward group was set on a fixed interval schedule to limit the possibility of participants making additional legal moves to receive more candy. Even though the reward group received candy for avoiding illegal moves these participants made the most illegal moves (M = 6.85). Such findings were unexpected and did not seem to be explainable by CIF. Some potential explanations could have been that the candy was not motivating enough or that the interval schedule was confusing or detrimental in some way. Participants only received one piece of candy for each successful 30 s time interval. It is possible that with more candy offered or a more substantial reward performance could have been improved. In addition, because the reward was administered on a fixed interval schedule participants could have had optimal performance that occurred early in the 30 s interval, but did not receive a reward until the end of the interval or potentially not at all if an illegal move was subsequently made in that interval. The lack of a significant reduction in illegal moves was surprising here, but the absence of such findings may have been due to a fault in the experimental design. One alternative explanation for the lack of significant reductions in illegal moves for the motivation and reward groups could have been that aversive consequences are necessary to reduce illegal moves.

The most surprising finding in Experiment 2 was the lack of a significant difference between the punishment and no-cost group. The punishment group was also set to administer a penalty at a fixed interval to maximize the similarity between the punishment and reward groups. However, this allowed participants to make illegal moves that did not result in a penalty because multiple illegal moves could have been made in one 30 s interval that only resulted in one penalty. This design may have reduced the effect of the punishment and was likely responsible for the absence of significant reductions in illegal moves.

Experiments 1 and 2 specifically looked at Stage 1 of CIF where consequences for making or avoiding illegal moves were manipulated in an attempt to reduce illegal moves. Experiment 1 lent support to CIF's claim that punishment and threat without punishment resulted in illegal move reductions. Experiment 2 indicated that the reward manipulation was not able to reduce illegal moves, but that a motivational manipulation showed promise. Further research is required to explore what other consequences, if any, are able to assist problem-solvers in avoiding illegal moves. According to CIF, manipulations in the first stage of the framework reduced illegal moves by acting on Stage 2 to increase a problem-solvers attention towards the problem. CIF made no specific assumptions about the attentional resources available and how they could be distributed to reduce illegal moves, only that problem-solvers do not often perform at optimal levels and that the consequences in Stage 1 would result in additional resources attributed to the task. Stage 2 of CIF was not specifically tested here and further research is required to determine how and where attention would be increased.

Stage 3 of CIF proposed that increased attention from Stage 2 would allow for increased accuracy and/or frequency of checking and/or evaluation. The checking portion of the framework utilized the Generation, Caution, Verification Framework (GCV) from Knowles and Delaney (2005). Experiment 3 specifically looked at checking behavior and its ability to reduce illegal moves. Participants in the experimental conditions were instructed to check each move for legality prior to executing that move, approximately half of those participants did this silently and the other half did so aloud. The check (M = 3.00), aloud (M = 2.00), and punishment (M = 2.65) groups made significantly fewer illegal moves compared to the no-cost group (M = 5.90). These results indicated that increased frequency of checking resulted in reduced illegal moves, lending support to that concept of the framework.

The check and aloud groups were both instructed to check moves for legality and to click a button on the display prior to each move to indicate that they had checked that move, this was done silently or aloud, respectively. There was a significant relationship between checking and illegal moves for the aloud group, but not for the check group, indicating that those in the aloud group reduced the number of illegal moves they made as checking increased. It was surprising that data from the check group did not yield a significant relationship, especially since performance in this group was similar to the aloud group where there was a significant relationship. One potential explanation could have been that since these participants did not check aloud they could have not been checking at all and could have just simply clicked the button prior to making a move. This explanation did not seem very plausible considering that participants in the threat group made a similar number of illegal moves compared to the aloud group, $t(38) = 1.24$, $p > .05$. Another, plausible explanation was that participants in the threat group were checking moves for legality in their head, but they were less likely to click the check button. This explanation also did not seem likely because an independent samples *t*-test between the two groups for number of moves not checked revealed no significant difference, $t < 1$. One final explanation could have been that both groups were checking and rejecting illegal moves, but the aloud group did so with more accuracy. Although this final explanation appears to be the most likely it is not possible to determine the true cause based on current findings.

Taken together the three experiments provide valuable information about the framework presented and about problem-solving behaviors more generally. Experiment 1 lent support to CIF's assumption that a physical penalty was not required to reduce illegal moves and that threat alone helped to induce caution. Experiment 2 presented somewhat confusing results and further research is required to determine what other types of consequences could reduce illegal moves. Experiment 3 directly looked at increasing checking behaviors and lent support to the claim that increasing the frequency of checking reduced illegal moves. These findings support CIF's claim that less intrusive manipulations can induce caution and that checking behaviors directly reduce illegal moves.

Transfer Effects

In all three experiments, each participant solved an *isomorph* of the first problem where the second problem had the same underlying structure and solution as the first problem, but did not share any outwardly similar features. This isomorph was solved after the first solution and was always solved under no-cost instructions. Having each participant solve the second problem with the same instructions and manipulations allowed for the observation of transfer effects from learning obtained on the first problem. For all comparisons, except one, the participants made a similar number of illegal moves on the second problems when compared with the first, which largely supported previous findings from Knowles and Delaney (2005). These findings seemed to indicate that the manipulations on the first solution, where illegal moves were reduced in most groups compared to the no-cost group, provided participants with sustained benefits that carried over to the second problem where illegal moves remained relatively low. However, there was one exception and some trends that should be addressed.

In Experiments 1 and 2, participants in the no-cost conditions demonstrated trends toward making fewer illegal moves on the second solution compared to the first. These trends may have potentially indicated that participants learned something that could be transferred to a novel isomorph or maybe it indicated that participants were "settling down" and were less nervous so performance improved, although not significantly. In Experiment 2, participants in the reward condition yielded numbers that approached significance, $t(19) = 2.08$, $p = .051$, potentially indicating that fewer illegal moves were made on the second problem ($M = 4.15$) compared to the first ($M = 6.85$). Why participants in the reward group made so many illegal moves on the first solution is unknown. These illegal moves commissions on the first solution were likely driving the trend toward fewer illegal moves on the second solution, but these results seemed counterintuitive given previous research where incentives helped to improve performance. Wieth and Burns (2006) found that the incentive to leave the experiment early resulted in improved performance on both incremental and insight problems. In addition, they found that the incentive resulted in further processing of the problem and increased participants' memory of the problem they solved. These findings support the transfer effects seen where participants showed a trend for reduced illegal moves on the second solution.

Finally, participants in the aloud group of Experiment 3 made significantly more illegal moves on the second solution ($M = 4.60$) compared to the first ($M = 2.00$), $t(19) = 2.73$, $p = .013$. This finding was unexpected and could have potentially been due to the aloud instructions causing the participants to take more time on the first solution. The participants may have felt the need to "make-up time" on the second solution, which could have led to carelessness and an increase in the number of illegal moves on the second solution. Another potential explanation could be that the aloud instructions prevented some necessary encoding or learning on the first solution, as transfer effects with aloud instructions were not explored in previous studies. The check group made a similar number of illegal moves on the first and second solution, so the checking instructions were not likely the cause of the increase. However, it was possible that the interaction between the think aloud and checking instructions could have been responsible for participants making more illegal moves on the second problem when compared to the first. This was a very interesting finding and the author believes that this warrants further exploration as the potential for think aloud instructions to affect, not performance, but learning could indicate a negative side effect of this manipulation not seen in previous research (Delaney, Ericsson, & Knowles, 2004; Knowles & Delaney, 2005).

Untested Assumptions: Illegal Move Reduction

CIF was proposed as a novel framework and therefore made several assumptions as to how a problem solver is able to reduce illegal moves. In the first stage of the framework the assumption was that various consequences for making illegal moves influenced the problem-solver to alter some aspect of his/her behavior. In the second stage the assumption was that the consequences in Stage 1 influenced the problem solver to increase the amount of attention devoted to the problem. With this increased attention the problem-solver then either increased the frequency and/or accuracy of either checking or evaluating behavior in Stage 3 of the framework. Because CIF was a novel framework it was not possible to test every assumption or every aspect of the framework here. Several assumptions remain untested and require future research to assess the ability of the framework to describe human problem-solver performance.

Experiments 1 and 2 addressed Stage 1 of the framework and attempted to assess what consequences would invoke caution and reduce illegal moves. Threat, punishment, motivation, and reward were all tested and threat and punishment successfully resulted in illegal move reductions while motivation showed promise. The reward consequence did not prove to reduce illegal moves. In addition, there may be other consequences not mentioned here that have the ability to reduce illegal moves and further research would be necessary to determine what, if any, other consequences would reduce illegal moves.

The assumption of CIF was that consequences in Stage 1 of the framework would motivate the problem-solver to devote more attention to the problem-solving episode. Previous research has indicated that illegal moves may occur due to resource limitations (Jeffries, Polson, Razran, & Atwood, 1977), recent research seemed to indicate otherwise (Knowles & Delaney, 2005). Because a resource limitation hypothesis cannot fully explain illegal moves, it made sense to assume that we do not always perform to our full potential and that we may have additional resources available to perform more efficiently. The framework assumed that these additional resources were represented as attentional focus and that consequences in Stage 1 tap into these additional attentional resources to improve performance. However, the framework made no assumptions as to how this additional attention was evoked or to how it was distributed, only that it exists and that it can be utilized, the current results seemed to support these assumptions.

Experiment 3 specifically looked at the third stage of the framework and how increased checking behavior would influence illegal moves. This increased checking resulted in reduced illegal moves as the framework predicted. However, changes in evaluation behavior and increasing the accuracy of checking were not assessed. It seemed intuitive that as the accuracy of checking a move for legality increased the likelihood of correctly rejecting illegal moves would also increase. However, further exploration is necessary to determine when and to what degree a problem-solver may increase the accuracy of checking candidate moves. CIF also assumed that as attention towards the problem increased a problem-solver may begin evaluating candidate moves more frequently or evaluating moves to a higher criterion.

Integration With Existing Theories of Illegal Move Selection

One of the main goals of this work was to address why illegal moves are made. One potential explanation for this, proposed by Jeffries and colleagues (1977), was that illegal moves are made due to resource limitations. Although the term "resources" was not well-defined, it usually refers to working or short-term memory limits, attentional capacity, and thinking speed. This explanation proposed that on problems such as those used in these experiments participants do not have the capacity to either remember to check a move for legality or to calculate the resulting state properly to assess legality. However, recent research has seemed to indicate otherwise. Knowles and Delaney (2005) found that when penalized for illegal moves participants were able to avoid illegal moves. The current work found similar results involving not only penalty, but the mere threat of a penalty indicating that problem-solvers do have the resources necessary to avoid illegal moves. In addition, work on various types of problems such as the 8-puzzle, the Tower of Hanoi, and water jugs problems has demonstrated that problem-solvers actually have the ability to plan out solutions to difficult problem states and entire problems themselves, again indicating that we have the resources necessary to perform better on these tasks (O'Hara & Payne, 1998; Welsh, Cicerello, Cuneo, & Brennan, 2001; Delaney, Ericsson, & Knowles, 2004).

Just because we can perform better on certain tasks than the resource theorists predicted does not imply that we have no limits on our cognitive resources. In fact, a central tenet of cognitive psychology is that many of our cognitive abilities are measurable and have documented limits. For example, there are significant limits on our short-term memory (Miller, 1956). Furthermore, Anderson and colleagues have successfully used 185 ms as an estimate of time to shift visual attention in versions of ACT-R (Anderson & Lebiere, 1998; Gray, Sims, Fu, & Schoelles, 2006). Because we have measurable cognitive limits, a resource limitation hypothesis cannot be entirely ruled out. However, on the tasks described above they likely play a smaller role than initially assumed, and resource theorists need to be more clear about exactly what resource is taxed and how.

Another possible explanation is that mental resources may not place a hard limit on our ability to check for illegal moves, but that we are "cognitive misers" who try to conserve resources as much as possible. Previous and current research has demonstrated through strategy changes, planning, and improved performance that we are often much more capable than our initial performance would indicate (for example Simon & Reed, 1976; O'Hara & Payne, 1998; Delaney, Ericsson, & Knowles, 2004; Knowles & Delaney, 2005). In Wilson's (2002) review article of embodied cognition, she stated that one of the views of embodied cognition was an off-loading of cognitive work onto the environment. This off-loading referred to the strategy of accessing information in the environment as needed instead of using resources, like committing such information to memory. In tasks such as the hobbits and orcs, problem-solvers may have used trial and error strategies instead of planning deeply or checking moves for legality, thereby conserving resources. A trial and error strategy would likely result in many illegal moves. However, with no penalty for illegal moves, such a strategy was not very costly in terms of resources, and was usually successful in advancing the problem-solver to the goal. Such an explanation seems to fit well with participants' performance in the no-cost groups, but it does not explain why a threat or penalty would change such behavior.

Gray and colleagues referred to this off-loading concept as the minimum memory hypothesis and compared this to the soft constraints hypothesis to describe participants' behavior on the Blocks World task (Gray et al., 2006). The *soft constraints hypothesis*, in contrast to the minimum memory hypothesis, stated that optimal performance was based on the currency of time and not on conserving memory resources. Gray et al. (2006) found that as the time cost of a task increased participants began to utilize their memory more to avoid procedures that took additional time, ultimately reducing the total time spent on the task. The idea of participants adjusting behavior to reduce the total time on a task actually fits well with the results observed in the first two experiments. Participants in the punishment group avoided illegal moves because the penalty phase automatically added an additional 45 s. Those in the threat and experience groups may have avoided illegal moves because they were told that illegal moves would have added additional time at the end of the task. It may have been possible that participants in the motivation group demonstrated a trend towards reduced illegal moves because they saw the motivating instructions as a hint to avoid illegal moves to reduce total solution time. Finally, participants in the reward group may not have shown illegal move reductions because doing so would not have decreased the time to solve the task. In fact, attempting to maximize the reward would have potentially taken longer to solve the problem.

The soft constraints hypothesis seemed to offer a valid explanation for many of the results obtained, in at least Experiments 1 and 2. Such findings offered a novel explanation for some of the unaddressed aspects of CIF. Stage 2 of CIF indicated that the consequences in Stage 1 resulted in an increase in attention. However, the soft constraints hypothesis may have indicated that increased attention was actually a reevaluation or redistribution of attention to reduce the total time a problem-solver spends on the task.

A reevaluation or redistribution of attention in the CIF framework could have been explained as a shift in strategy. The author believed that participants initially engaged in a trial and error strategy because there was little or no cost for illegal moves and this strategy was often successful at eventually reaching the goal state. Increasing the cost of making illegal moves may have lead to a representational change of the problem and a shift in strategy. Lovett and Schunn's (1999) RCCL model predicted that unsuccessful strategies would lead to a representation change of the problem and through learning unsuccessful strategies were abandoned and successful strategies were adopted. Simon and Reed (1976) also found a strategy shift towards means-ends analysis when participants were given a hint of a state they would encounter during their solution. Because illegal moves were highlighted in the threat and punishment groups the execution of such a move may have seemed like a failed strategy to the participants and prompted a representational change or a strategy shift. This idea seemed plausible as adopting a more successful strategy could have likely resulted in illegal move reductions.

An additional explanation for the avoidance of illegal moves could have been attributed to problem-solvers "learning" from their illegal moves. Grobe and Renkl (2007) had participants begin by working on pretest probability problems to assess their prior topic knowledge. Next they presented participants with worked examples of probability problems. Some of the participants were presented with worked solutions that were all correct and some participants were presented with a mixture of correct and incorrect solutions (participants were told when a solution was incorrect). Following this exercise, participants completed a post-test of problems that varied in their similarity to the worked examples they had seen. The post-test results indicated that those participants with low prior knowledge benefited from seeing the correct solutions only, but those participants with high prior knowledge benefited from seeing both the correct and incorrect solutions.

In the current experiments participants were not assessed on their prior problem-solving knowledge, but it may have been possible that at least some of the participants were able to learn from the illegal moves they made. Then why were illegal moves reduced in the experimental groups compared to the no-cost control? Threatening or penalizing participants for illegal moves may have increased their attention to the task and motivated them to learn from their illegal moves so that they could be avoided in the future. With little or no cost to making illegal moves learning may have seemed like a waste of time or waste of resources. After participants in the experimental groups made illegal moves they may have been more likely to assess why that move was illegal and this would have aided in avoiding illegal moves in the future. This would explain why transfer effects to novel isomorphs were seen where participants continued to make a reduced number of illegal moves. A problem-solver's ability to learn from his/her illegal moves seems like a plausible explanation not only for transfer effects, but for illegal move reductions on the initial problem where the first illegal move has the potential to provide a valuable lesson.

Learning may have occurred from reflecting on an illegal move, but it may have also been possible that learning took place during the penalty phase. Moss, Kotovsky, and Cagan (2007) presented participants with a Remote Associates Test (RAT) where participants had to find a word that was associated to a group of words presented. After several RAT problems, participants completed a lexical decision task that provided hints to the RAT problems that were not solved correctly. Participants were then returned to the previously solved and unsolved RAT problems they had attempted. The results indicated that the hint appeared to help participants solve the uncompleted RAT problems. One explanation Moss et al. considered was that of an incubation period. What actually occurs during an *incubation* period is not fully known, but it is essentially a "break" from the problem where the problem-solver may continue working on the problem subconsciously or the break may influence a strategy change. One example of where an incubation period is often helpful is with insight problems where the problem-solver can often return to the problem with greater success on a subsequent attempt. It could be that in the current experiments the 45 s penalty phase acted as an incubation period where participants either continued working on the problem or this may have prompted a strategy change. This explanation seemed unlikely as their was no incubation period in the threat or experience groups in Experiment 1, which both demonstrated illegal move reductions compared to the no-cost group.

Previous theories from the problem-solving literature and other areas of research help to provide alternative explanations and insight to the current findings. Although resource limitations may play a role in the current findings their influence likely plays a smaller role than initially assumed. Resource limitations may have seemed like a valid explanation for many results because problem-solvers may be "cognitive misers" and conserve resources or they may have an alternative goal of conserving time. It is through experimental manipulations that we are able to gain insight into the true potential of the human problem-solver. When there are changes to the problem-solving environment, as with punishment or threat, we see that the problem-solver adapts. This adaptation could be defined as an attentional shift or strategy shift, but no matter what the term problem-solvers are able to alter performance and improve. The current findings contribute to the problem-solving literature and provide additional insight into problem-solvers' ability to adapt and improve.

How Do People Evaluate Moves?

Several move evaluation functions have been proposed in the problem-solving literature. However, it did not seem wise to adopt one of these functions to CIF without first testing to see which one had the best fit. At the same time, illegal moves are most likely selected during the move selection phase and only rejected during evaluation or checking. Therefore, it is worth considering how earlier theories have accounted for move selection, with an eye towards understanding why illegal moves are chosen according to those theories. One of the very first theories of problem-solving that looked specifically at how problem-solvers solved a problem and what they were doing was Newell and Simon's (1988)[1] General Problem Solver (GPS) program. *GPS* was a computer program that attempted to mimic human problem-solving performance by creating and completing subgoals to reach the ultimate goal of advancing to the final goal state. Newell and Simon had a participant think aloud as he/she worked on a symbolic logic problem and then used the verbal protocol to compare performance to the computer program's attempt at solving the problem. Although, GPS did not map the participant's moves exactly the findings proved promising and have been influential to problem-solving research.

[1] Newell and Simon's 1988 work was read by the author as a chapter reprint. GPS was previously introduced by Newell, Shaw, and Simon in 1959.

Since the introduction of GPS back in the late 1950's models of human problem-solving have evolved and become more specified. In fact, a model of human problem-solving that was discussed in the first chapter was proposed specifically for problems like the hobbits and orcs. The model proposed by Jeffries et al. (1977) included the evaluation function:

$$e_i = aM_i + bC_i + cP_i \qquad (1\text{-}1).$$

The assumption of this evaluation function, as with most evaluation functions, was that a problem-solver was able to assign candidate moves a value and determine which move had the highest probability of advancing the problem-solver towards the goal state. Due to the design of the evaluation function those moves that placed more travelers on the goal side (the right bank) of the river and did so by maintaining a balance of missionaries and cannibals (missionaries are interchangeable with hobbits and cannibals with orcs) to reduce the probability of having a missionary eaten would be more likely to be selected. Such a strategy seemed intuitive and potentially a good fit for CIF.

The assumptions of CIF stated that illegal moves may be reduced through increasing the frequency and/or accuracy of evaluating moves. Increasing the frequency of evaluating moves would simply entail using an evaluation function, like this one, more often instead of selecting moves at random. One argument was that by increasing the frequency of evaluation the probability of selecting an illegal move actually increased because many illegal moves actually bring the problem-solver closer to the goal and would therefore be evaluated higher than legal moves. The evaluation function proposed by Jeffries et al. indicated that problem-solvers attempt to keep the travelers in missionary-cannibal pairs to minimize the chance of the missionaries being eaten. This assumption of the evaluation function would help to prevent illegal moves sometimes, but the problem cannot be solved unless the pairs are split. Since a split of the pairs was inevitable the threat of evaluation increasing illegal moves remained. Jeffries et al.'s model also stated that there were two types of illegal moves, easy-to-detect and hard-to-detect and that easy-to-detect illegal moves were always rejected. Hard-to-detect illegal moves were checked and rejected based on fixed probabilities. CIF assumptions differed here in that these probabilities were not fixed and that checking could occur independent or as a results of evaluation. If the evaluation of moves can occur in CIF without checking behavior then the possibility of illegal moves increasing with increased evaluation cannot be ruled out.

The second alternative for evaluation to aid in avoiding illegal moves according to CIF was by increasing the accuracy of evaluation, which can be done in two different ways. One way was by including an illegal move filter as part of the evaluation function to reject illegal moves as noted previously. The second assumed that problem-solvers are affected by different consequences that alter performance. These consequences increase attention towards the problem, which may potentially increase learning and the integrating of past experiences to increase the probability of selecting legal moves more often. In their model, Jeffries et al. included a memory process that allowed for previously visited states to be avoided to prevent backtracking and to facilitate progression through the problem, but since it was based on the resource limitation hypothesis little or no learning or adaptation to the problem states were incorporated. Based on this it would not be likely or maybe not possible that a problem-solver could solve the problem in the minimum number of moves without any illegal moves. However, out of the 240 participants that completed both problems 9 were able to solve the problem in the minimum number of moves with no illegal moves in the current experiments. These participants for the most part were in various groups and in different experiments.

To help explain how a problem-solver is able to efficiently solve this problem with no wasted moves a brief summary of the solution of the hobbits and orcs problem is provided here (the solution is almost completely linear so all participants solved the problem in this order). First the problem-solver had to position all of the orcs on the right bank of the river. Next the hobbits were moved over in a balanced fashion so that there were two hobbits and two orcs on the right bank. Then for the first and only time in the problem the problem-solver moved two travelers to the left bank, one hobbit and one orc. At this point the problem solver could safely move all the hobbits to the right (goal) bank where they remained until the rest of the travelers were moved to the goal bank. To utilize the current evaluation function to explain this performance the constant weighting factors a, b, and c would need to be adjusted at different states of the problem. For instance, b, which accompanied the cannibals' value on the right bank would need to be weighted heavier towards the beginning of the problem to ensure that the cannibals were moved first. Once this phase was completed the weighting factors again would need to be adjusted so that a, which accompanied the missionaries, was weighted more heavily to ensure that they would be moved. Increasing the accuracy of the evaluation function would require that participants adjusted these weighting factors at each state and this increased accuracy would likely aid in the avoidance of illegal moves.

The evaluation function and model proposed by Jeffries et al. (1977) appeared to provide a good description of participants' problem-solving behavior presented here with minor revisions. The major assumptions of their model were based on a resource limitation hypothesis, but since the current and recent research seemed to indicate otherwise such adjustments seemed warranted. Because this evaluation function was based on the same types of problems used here it had a natural fit into the CIF framework. However, it would a gross oversight to think that research and models taken from other types of problems could not provide valuable insight as well.

An additional model that could be used to explain the results obtained here was based on research that has utilized John Anderson's ACT* theory (often referred to now as ACT-R). ACT* was based on the firing of *productions*, which were condition-action pairs (Anderson, 1987). If a *condition* was satisfied then the corresponding *action* would be executed, e.g. if the traffic signal is red, then I will stop my vehicle. Lovett's (1998) work on the building sticks task (BST) has utilized the ACT-R theory to model problem-solvers' ability to choose moves that will likely reach the goal state. Lovett's ACT-R based model assumed that problem-solvers selected their next move based on the highest expected gain according to the following equation:

$$E = PG - C \tag{1-2}$$

where E was the expected gain of the selected move, P was the estimated probability of achieving the production's goal, G was the value of the goal, and C was the estimated cost to be expended in reaching the goal.

Since previous research indicated that problem-solvers have the resources to calculate future states and to plan then they should have the ability to calculate expected gains (E) that are equal to true gains in tasks such as the hobbits and orcs. Unlike BST, the hobbits and orcs task was not solved several times and it had a fixed solution and fixed goal state that was not manipulated by the experimenter. The problem solver had to advance through the problem space encountering many novel states along the solution path. At any given state subsequent states could have been calculated and evaluated as to their ability to advance the participant to a novel legal state and the success of an operator was independent of its history of success. In addition, the problem space for the hobbits and orcs problem was almost completely linear. This enabled participants to solve the problem and completely avoid all illegal moves by correctly calculating future states and only selecting novel legal moves. In contrast, it was not possible to accurately calculate the validity of a move in BST, influencing participants to rely primarily on history of successes and failures of operators in these types of problems. Lovett's model, as applied to BST, focused primarily on participants' ability to make decisions based on previous successes and failures. However, with some modification it may be possible to apply Lovett's model to the hobbits and orcs problems to account for illegal move reductions obtained in previous and current research.

In Lovett's model P was the estimated probability of achieving the production's goal and was made up of the product of qr. The probability that the production would have achieved the desired state q was fixed to 1 and the probability of achieving the production goal given arrival at the intended state r was based on previous successes and failures. However, the hobbits and orcs task differs greatly from the BST in that previous successes and failures have less of an impact and problem-solvers could have calculated future states and illegal moves based on the rules. To adapt this model to the hobbits and orcs problem r would have been fixed to 1 because successes and failures contribute little to the subsequent selection of operators and q would = the problem-solvers' ability to correctly calculate the future state. It was possible that q would also contain the 3-stage framework GCV proposed by Knowles and Delaney (2005) if the move were to be assessed for legality. However, previous research seemed to indicate that when the cost for making illegal moves was low problem-solvers often did not engage in this activity. With an increase in cost for an illegal move problem-solvers would have likely increased the frequency and/or accuracy with which they engaged in calculating q resulting in P yielding higher values for legal moves and the reduction of illegal moves.

As indicated above, move evaluation functions from other problems or from similar problems may have the potential of being updated and integrated into the CIF framework. However, additional research is required to determine if a move evaluation function can be adapted from another problem, from similar problems, or if a novel move evaluation function would provide the best fit for CIF.

Real World Application and Future Research

Problems are something that we encounter every day. Whether it is balancing our checkbooks, driving to work, or cooking a meal, these daily tasks consume a large portion of our lives and understanding how to solve these problems more efficiently would offer great benefits. Laboratory results have revealed that one way to improve performance may be through punishment (Knowles & Delaney, 2005). However, being punished for every mistake or illegal move we make does not sound like a lot of fun. Therefore finding alternative ways to improve performance seemed desirable.

The current work looked at alternative ways for improving problem solving and discovered that a threat without punishment reduced illegal moves. Motivation and reward did not result in significant reductions, but motivation showed promise. These findings seemed to support some things that we already do or experience today. You threaten your child not to touch the stove for he will get burned or your boss threatens you that you will get fired if you show up late for work again. The findings presented here indicated that a credible threat can yield improvements as significant as punishment, at least on these types of problems. The motivation and reward results were somewhat surprising based on common practices in the world today. Motivational seminars and tapes are utilized by companies and individuals in an attempt to enhance performance and rewards are offered in the form of money or other incentives for meeting or exceeding expected goals. Because motivation and rewards are successfully used in our society to drive performance it strengthens the claim that the design lacked the power to detect these effects or that the motivating instructions or rewards were not adequate. So then what has this work offered?

The CIF framework made several assumptions about how we solve problems and operate in a given problem space. The main assumption of CIF was that we do not perform optimally in certain situations and that we have the ability to do better. The results presented here supported that assumption by demonstrating that with consequences for illegal moves or by increasing cautiousness through checking we can reduce illegal moves and increase problem solving efficiency. One potential method for avoiding illegal moves is by being more cautious and checking moves for legality. We can liken this to an airline pilot who has a checklist of items he must verify and complete before he is able to take-off. Of course, we do not need a checklist for everything we do in life, but being more cautious in certain situations would serve us well. Some examples of areas in life where we could apply this information and benefit include: driving (checking and abiding by the speed limit and other laws would likely prevent us from getting a ticket or getting into an accident), finances (checking our bank balance before writing a check to ensure that we do not overdraw our account), and military training (ensuring that a soldier knows the proper procedures for a critical situation so that no one gets injured). The real world applications and future avenues for such findings seem limitless as we are constantly striving to become more efficient to conserve time and/or resources.

The results obtained here seemed to support the CIF framework that certain consequences of illegal moves can reduce illegal moves and that checking moves for legality may be one method for reducing those moves. However, several questions still remain and future research should focus on these questions. First, what other types of consequences, including different types of motivators and rewards, are able to reduce illegal moves? Second, Stage 2 of the framework should be further studied to determine if an attentional shift is responsible for increased caution and if so how this attentional shift reallocates these additional resources. Third, future research should also focus on the evaluation function used by problem-solvers in determining which move to select next in the problem-solving episode. Finally, how the accuracy of checking is altered over the duration of a problem should be explored to determine the likelihood of errors being made after a problem-solver has decided to check a move for legality.

CHAPTER 6
CONCLUSION

Three experiments were presented to assess the ability of the CIF framework to describe and predict problem solving performance and the reduction of illegal moves. The main assumption of the framework was that consequences of illegal moves increased attention to reduce illegal moves. The first experiment tested whether a penalty was required during the problem solving phase to increase attention or whether the threat of penalty alone would be sufficient to increase attention and reduce illegal moves. The second experiment addressed whether an adverse stimulus or threat of an adverse stimulus was required or whether motivation or reward could be effective at inducing caution and increasing attention. Finally, the third experiment directly addressed a portion of the framework by instructing participants to check moves for legality and also collected verbal protocols to assess what participants were actually doing/thinking.

Experiment 1 results indicated that punishment was not required to increase attention as significant results were found in both the threat and experience conditions. These findings demonstrated that less intrusive interventions could potentially be as effective in increasing attention when compared to punishment that occurred during the problem solving episode. The results of Experiment 1 have significant real world applications that could be helpful when punishment can have serious negative effects or when punishment is not feasible during a problem solving episode. However, further research is required to determine the true applicability of these results to real world settings.

Experiment 2 did not achieve the predicted results and the motivation and reward conditions were not effective at reducing illegal moves. The lack of illegal move reductions could be attributed to one of several explanations. Experiment 2 utilized undergraduate participants and was run at the end of the semester. Thus, the participants in this experiment were likely those that waited until the end of the semester and may have been less motivated and resistant to the experimental manipulations. Alternatively, the motivating instructions and the candy reward may have not been sufficient to increase attention and a different motivation or more valuable reward may have achieved significant reductions. Finally, it was also possible that there were no “end of semester” effects and there was nothing wrong with the motivating instructions or reward and that an adverse stimulus or threat of adverse stimulus was necessary to increase attention. Experiment 1 demonstrated that a punishment was not necessary to reduce illegal moves, but the concern of future punishment was able to reduce illegal moves. The concern of future punishment still had the belief of something negative and may be a necessity in inducing caution.

Experiment 3 directly manipulated and tested the third stage of the CIF framework. Instructing participants to check moves for legality reduced the number of illegal moves made whether participants were instructed to think aloud or not. This manipulation was the first that did not focus on the consequences of making an illegal move and attempted to reduce illegal moves by avoided them before they were made. This direct manipulation revealed that increasing the frequency of checking resulted in a decrease in illegal moves. This finding also has practical applications in that recommending or requiring a person to check for legality in problem solving episodes may help increase problem solving efficiency by reducing illegal moves.

One unexpected result obtained in Experiment 3 that was not addressed was the increase in illegal moves from the first to the second problem for the aloud group. This release from benefits was likely due to an interaction between the thinking aloud and checking as there was no release found in the check group and no release in aloud groups in previous research (Knowles & Delaney, 2005). Further investigation is required to determine the specific cause of the release.

In conclusion two experiments lent support to the CIF framework and indicated that increased attention was a likely cause of reductions in illegal moves. The threat of punishment and instructions to check moves for legality both reduced illegal moves and provided support to the framework. However, reductions in illegal moves were not discovered with motivating instructions or reward, possibly indicating that an adverse stimulus or threat of adverse stimulus was required to induce caution and increase attention. In conclusion, these experiments provided support for CIF, with the exception of Experiment 2, and indicated that the framework may be a valuable tool in determining potential behaviors on such tasks. Additional research is required to further explore the ability of CIF to predict problem solving behavior and to determine if rewards and motivation have the ability to reduce illegal moves.

APPENDIX
THINKING ALOUD INSTRUCTIONS: EXPERIMENT 3

The thinking aloud instructions presented to a participant in Experiment 3 were as follows: "In this experiment you will be asked to solve problems on the computer using the mouse. However, as you work on the problem you will be asked to think out loud. That is, as you work on the problem I would like you to say whatever comes into you head. It doesn't matter if it makes sense or not, I just want you to say whatever you are thinking. For practice, try this simple task: Think of as house that you are very familiar with, it could be your's or a friend's. Think out loud as you imagine yourself walking through the house counting the number of windows you see. You may begin whenever you are ready."

LIST OF REFERENCES

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review, 94(2)*, 192-210.

Anderson, J. R., & Lebiere, C. (Eds.). (1998). *Atomic components of thought.* Hillsdale, NJ: Erlbaum.

Atwood, M. E., & Polson, P. G. (1976). A process model for water jug problems. *Cognitive Psychology*, *8*, 191-216.

Delaney, P. F., Ericsson, K. A., & Knowles, M. E. (2004). Immediate and sustained effects of planning in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1219-1234.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Gray, W. D., Sims, C. R., Fu, W. T., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review*, *113 (3),* 461-482.

Grobe, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction, 17*, 612-634.

Jeffries, R., Polson, P. G., Razran, L., & Atwood, M. E. (1977). A process model for missionaries-cannibals and other river-crossing problems. *Cognitive Psychology*, *9*, 412-440.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled attention view of working memory capacity. *Journal of Experimental Psychology: General*, *130*, 169-183.

Knowles, M. E., and Delaney, P. F. (2005). Lasting reductions in illegal moves following an increase in their cost: Evidence from river-crossing problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 670-682.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, *17*, 248-294.

Kotovsky, K., & Simon, H. A. (1990). What makes some problems really hard: Explorations in the problem space of difficulty. *Cognitive Psychology*, *22*, 143-183.

Lovett, M. C. (1998). Choice. In J. R. Anderson & C. Lebiere (Eds.), The atomic components of thought (pp. 255-296). Mahwah: Lawrence Erlbaum Associates.

Lovett, M. C., & Anderson, J. R. (1996). History of success and current context in problem solving: Combined influences on operator selection. *Cognitive Psychology, 31*, 168-217.

Lovett, M. C., & Schunn, C. D. (1999). Task representations, strategy variability, and base-rate neglect. *Journal of Experimental Psychology: General, 128(2)*, 107-130.

Luchins, A. S. (1942). Mechanization in problem solving. *Psychological Monographs, 54* (Whole No. 248).

Miller, G. A. (1956). The magical number 7, plus or minus 2: Some limits on our capacity for processing information. *Psychological Review, 63(2),* 81-97.

Moss, J., Kotovsky, K., & Cagan, J. (2007). The influence of open goals on the acquisition of problem-relevant information. *Journal of Experimental Psychology: Human Learning and Memory, 33(5)*, 876-891.

Newell, A., & Simon, H. A. (1988). GPS, a program that simulates human thought. In A. M. Collins, E. E. Smith, (Eds.), *Readings in cognitive science: A perspective from psychology and artificial intelligence* (pp. 453-460). San Mateo, CA: Morgan Kaufmann, Inc.

O'Hara, K. P., & Payne, S. J. (1998). The effects of operator implementation cost on planfulness of problem solving and learning. *Cognitive Psychology*, *35*, 34-70.

Platt, R. D., & Griggs, R. A. (1993). Facilitation in the abstract selection task: The effects of attentional and instructional factors. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 46A(4),* 591-613.

Simon, H. A., & Reed, S. K. (1976). Modeling strategy shifts in a problem-solving task. *Cognitive Psychology, 8*, 86-97.

Turner, M. L., & Engle, R. W. (1989). Is working-memory capacity task dependent? *Journal of Memory and Language*, *28*, 127-154.

Welsch, M., Cicerello, A., Cuneo, K., & Brennan, M. (2001). Error and temporal patterns in Tower of Hanoi performance: Cognitive mechanisms and individual differences. *The Journal of General Psychology, 122(1)*, 69-81.

Wieth, M., & Burns, B. D. (2006). Incentives improve performance on both incremental and insight problem solving. The *Quarterly Journal of Experimental Psychology, 59(8)*, 1378-1394

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9 (4),* 625-636.

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science*, *18*, 87-122.

BIOGRAPHICAL SKETCH

Martin E. Knowles was born in Huntington, New York, to James and Joyce Knowles in 1978. At the age of seven, he and his family moved to Tampa, Florida, where he lived until he was 18. At this time, he began attending Florida State University in Tallahassee, Florida, where he completed a Bachelor of Science degree in the fall of 2000. In 2001, Martin moved to Gainesville, Florida and began his graduate studies in cognitive psychology at the University of Florida, under the supervision of Dr. Peter F. Delaney. He married his wife Stephanie in the spring of 2004 and completed his Master of Science degree that same year. His master's work was subsequently accepted for publication in the Journal of Experimental Psychology: Learning, Memory, and Cognition. In March of 2006, Martin became a father to a healthy baby boy and subsequently moved to Jacksonville, Florida with his wife and son. He successfully defended his dissertation in August of 2008, and received his Doctor of Philosophy degree in December of 2008, from the University of Florida.