Lasting Reductions in Illegal Moves Following an Increase in Their Cost : Evidence From River-Crossing[*]

By: Martin E. Knowles and Peter F. Delaney

Knowles, M. E., & Delaney, P. F. (2005). Lasting reductions in illegal moves following increasing their cost: Evidence from river-crossing problems. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 31,* 670-682. Digital Object Identifier: 10.1037/0278-7393.31.4.670

Abstract:

Improving problem solving requires understanding what difficulties people have when they approach novel problems. Some of the known difficulties include identifying and understanding the operators (Kotovsky & Simon, 1990) or having implicit but wrong constraints on move selection (Richard, Poitrenaud, & Tijus, 1993), memory load imposed by the task or external circumstances (Kotovsky, Hayes, & Simon, 1985), whether the rules agree with real-world knowledge (Griggs & Cox, 1983; Kotovsky & Simon, 1990), difficulties in planning ahead (Atwood, Masson, & Polson, 1980; Atwood & Polson, 1976; Delaney, Ericsson, & Knowles, 2004), failures to sufficiently reflect on move selection (Davies, 2000), and heuristic biases that may lead problem solvers to select the wrong move at any given time (e.g., Atwood et al., 1980; Atwood & Polson, 1976).

Article:

Legal moves are those performed in compliance with the rules of the valid problem space, which consists of nodes representing each of the valid (legal) states of the problem—that is, the valid problem space is a map of all the legal problem states and their connections (Newell & Simon, 1972). These valid states are defined by the rules of the problem. Illegal moves are those moves that violate a rule and take the problem solver outside of the problem space. The research cited in the preceding paragraph generally identified difficulties with legal moves, but problem solvers often make many illegal moves as well.

Why study illegal moves and their contribution to problem difficulty? Selecting illegal moves always leads to wasted effort because such moves could never be part of a correct solution. Furthermore, illegal moves may result in serious consequences, as in the case of air traffic controllers who may instruct two planes to travel in the same airspace simultaneously. In other cases illegal moves may result in less serious consequences but nonetheless may constitute a large portion of problem-solving activity.

Despite the prevalence and potential importance of illegal moves, few researchers have focused on their reduction. Zhang and Norman (1994) found that how the rules of the problem were represented greatly influenced the number of illegal moves committed on isomorphs of the Tower of Hanoi problem. Zhang and Norman drew a distinction between *internal* rules, which need to be explicitly stated in the problem description and then retained in memory by the problem solver, and *external* rules, which are enforced by the physical environment and therefore need not be explicitly stated or retained in memory. External rules resulted in

---

[*] Acknowledgement: The first experiment from this article was part of Martin E. Knowles's master's thesis (Knowles, 2004). We thank Chris Cardani, Danae Dunkley, Dania Hadjez, Alex Hepner, Travis Johnson, Michael Kung, Kevin McDonald, Jay Murray, Paul Niesen, Nathaniel Patterson, Karissa Plusko, and Mikal Rasheed for their help in collecting the data. Our thanks also go to Keith Berg and Ira Fischler, who served on Martin E. Knowles's thesis committee and who offered invaluable suggestions and feedback to improve this work, and to Christine Macdonald, who provided feedback on a draft.

significantly fewer illegal moves when compared with internal rules and also improved other solution characteristics such as solution time and solution length.

Kotovsky and Simon (1990) found that the type of information available to a problem solver and problem representation influence problem difficulty. They presented participants with all of the legal move options at each state during the solution of an isomorph of the Chinese ring puzzle. This manipulation eliminated participants' need to check each move for legality before selecting a given move and resulted in fewer illegal moves, less backtracking, and a decrease in solution time and length.

These works shed some light on how illegal moves can be reduced: They indicate that converting internal rules to external rules and presenting all legal options at any given state can help to reduce illegal move making. However, externalizing rules or presenting all legal options are not always possible. In addition, these experiments do not provide much insight into why illegal moves are made.

One of the few attempts to explicitly address the causes of illegal move selection focuses on resource limitations. Jeffries, Polson, Razran, and Atwood (1977) argued that when problem solvers are performing at or near their resource limits, they may miscalculate a future state or they may fail to check the future state for legality, resulting in an illegal move. Resource limitations provide an intuitive and plausible explanation for illegal move commission. Changing the rules from internal to external, as suggested by Zhang and Norman (1994), would tend to reduce memory load, consequently freeing up resources and resulting in fewer illegal moves. Similarly, presenting all of the legal alternatives, as suggested by Kotovsky and Simon (1990), would reduce the need for calculating future states and free up resources, resulting in fewer illegal moves. Resource limits can also explain the commonsense finding that with practice illegal moves tend to become rarer because as people gain more experience with a task, they find better ways of solving the problem, including relying on memory for previous decisions, more efficient search strategies, and so on. Such changes would tend to reduce the resource demands of the task and consequently reduce illegal move rates with practice.

In this article, we focus on identifying when illegal moves could be generated (or avoided). We propose three possible stages at which illegal moves might be generated. First, an illegal move must come to mind and be considered (illegal moves may not come to mind because the correct move can be retrieved from memory or because other options are more salient). If an illegal move does come to mind, then participants must remember to check the rules. If they check the rules, they must be successful in doing so in order to correctly reject the illegal move. Taken together, these three stages represent a metacognitive framework for investigating illegal move generation in problem-solving tasks. We apply the framework specifically to the hobbits and orcs river-crossing problem because it has been carefully studied before and, despite being unfamiliar to most people, it is usually eventually solved. The framework can be equally well applied to any problem-solving environment, however.

This framework was developed on the basis of a mathematical model proposed by Jeffries et al. (1977) for river-crossing problems. Jeffries et al. described two of these three stages as having fixed probabilities, where ε_1 was the probability of checking a move for legality and ε_2 was the probability of correctly rejecting an illegal move. ε_2 was assumed to be relevant to hard-to-detect moves, in which the orcs outnumbered the hobbits by only one on the bank opposite from the boat (see *Problem and Interface*). All other moves were easy-to-detect, and it was assumed that they were always checked and correctly rejected. Jeffries et al. assumed that ε_1 and ε_2 had fixed probabilities and that resource limitations were the major cause of the selection of illegal moves. Our framework makes no such assumptions and does not necessarily differentiate hard-to-detect and easy-to-detect moves.

Our three-stage framework is shown as Figure 1 and suggests ways to study the source of the impact of particular manipulations that reduce illegal moves. For example, if a manipulation (e.g., practice) reduces illegal moves, then one can attempt to identify the source(s) of the observed practice effect. Specifically, a given manipulation might produce its impact at any or all of the three stages. Because the framework is hierarchical, changes at the upper levels alter the magnitude of effects at lower levels by increasing or reducing the number of opportunities those lower level effects have to work. For example, if participants become very good at making legal moves, then presumably no illegal moves would even come to mind, which would stop most of the processing at the top level of the framework. This would mean that effects at lower levels would be difficult to detect. Therefore, when applying the framework one must either experimentally or statistically control performance at higher levels of the framework before lower levels can be reliably evaluated.

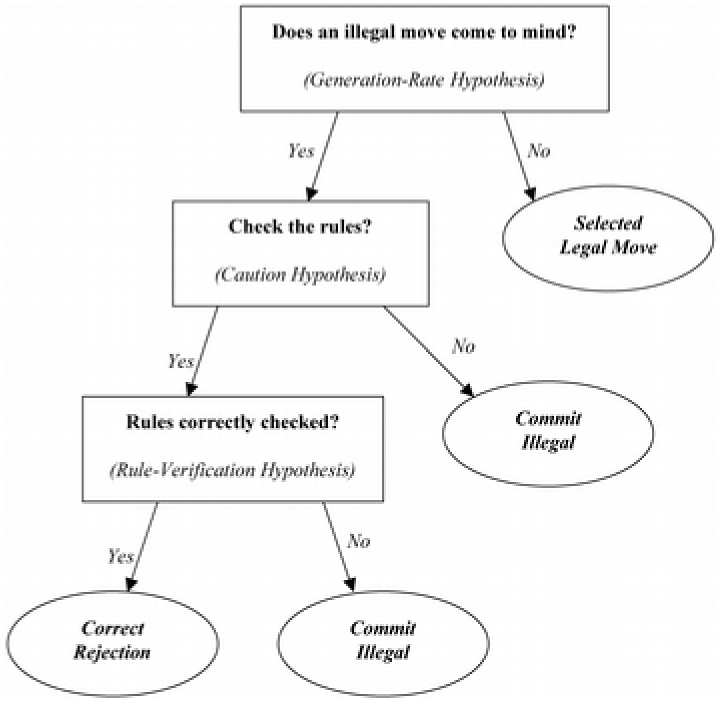

Figure 1. Three-stage framework for the selection and correct rejection of an illegal move

The first stage suggests a *generation-rate hypothesis*, which states that the likelihood of making an illegal move depends on the number of times participants make a decision. If the number of decisions can be reduced in some way, then the number of opportunities to generate an illegal move for consideration will also be reduced. This could occur if, for example, the order in which moves come to mind changes, which is known to occur in well-learned tasks such as chess (Saariluoma, 1992), or if the correct move can be retrieved directly from memory of previous solutions. Manipulations that make the problem “easier” by reducing legal moves would thus presumably have their impact on the generation rate. The generation-rate hypothesis generally predicts a positive correlation between the number of legal moves taken and the number of illegal moves made on a given problem.

The second stage gives rise to what we call the *caution hypothesis*, which proposes that a problem solver can become more likely to check candidate moves for legality over time. We find this hypothesis the most intuitively valuable to test because a manipulation that increases caution might transfer even to novel problems, as it likely reflects a general tendency to check moves for legality rather than any task-specific skill.

Finally, a third possibility, which we refer to as the *rule-verification hypothesis*, is that participants become increasingly effective at testing moves for legality with experience. This would imply that illegal moves that are considered would tend to be rejected more frequently with experience. The result would be fewer illegal moves even if the number of illegal moves that come to mind remains relatively fixed over time.

In the current work, we applied this framework to understanding one manipulation that we found reduces illegal move rates, namely, introducing a penalty (cost) for making illegal moves. To preview our results, we found support for the generation-rate hypothesis and the caution hypothesis. The cost manipulation reduced illegal move rates without affecting legal move rates (Experiments 1 and 2) without manipulating the amount of resources required to solve the problem. Even once participants were no longer penalized for making illegal moves, the cost manipulation continued to produce lasting reductions in illegal moves on a second solution of the problem (Experiment 2). The impact of the cost manipulation also transferred to novel problem-solving tasks even once the penalty for illegal moves was no longer enforced (Experiment 3). Taken together, the results suggest that the cost manipulation primarily increases cautiousness and that there are simpler ways than manipulating resources to reduce illegal moves during problem solving.

Although our primary purpose was to explore the cost manipulation, we also applied the framework to practice effects while replicating some classic experiments by Reed, Ernst, and Banerji (1974). Experiment 2 showed that with practice, participants make fewer illegal moves, although legal moves are not necessarily affected. Practice effects did not seem to be due to cautiousness (Experiment 3) and perhaps could be attributed to better learning of specific rules.

Experiment 1

We were mainly interested in whether introducing a penalty for making illegal moves would reduce their frequency and, if so, why. We introduced a manipulation whereby illegal moves were penalized with a time-out, and we compared participants' performance with performance by a control group that had no such penalty. If this cost manipulation reduced the rate of illegal move making, it would imply that directing participants' attention toward avoiding illegal moves could result in more efficient problem-solving behavior. If not, it would support a resource-based explanation whereby illegal move making cannot be easily unlearned. In addition, participants were instructed to work on the problem either silently or while thinking aloud. The thinking aloud data allowed us to calculate whether and when participants considered making an illegal move but correctly rejected the move before executing it.

Method

Participants

Participants were 70 undergraduates from the University of Florida who received course credit for their participation. Six participants proved unable to solve the problem within the allotted time of 20 min. They were assisted in completion of the problem and were therefore excluded; thus, 64 participants were included in the analyses.

Problem and Interface

The hobbits and orcs problem (which is 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 near the middle of the computer screen, with the boat beginning 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.

The problem was written in Microsoft Visual Basic and presented on a Gateway desktop, and 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 would remove them from the boat and place them back on the bank in the middle of the screen. If participants added too many travelers to the boat, allowed the orcs to outnumber the hobbits, attempted to move the boat with no travelers selected, or violated the rules in any other way they were notified via a message box and the move did not occur.

Design

The experiment had a 2 × 2 design: Verbalization (silent vs. aloud) × Cost (cost vs. control). Both variables were between-subjects.

Verbalization variable

Participants were randomly assigned to work on the problem either silently or while thinking aloud. In the think aloud condition, before starting the problem, participants were given instructions for thinking aloud similar to those suggested by Ericsson and Simon (1993). They then practiced thinking aloud by imagining that they were leaving school for the day, and they were to describe the path they were taking home and everything they saw along the way. In addition, participants were asked to state each traveler they were considering moving at each state of the problem even if they were not intending to execute that particular move. These additional instructions were given in an attempt to motivate the participants to produce more complete and elaborate verbal protocols. We are not aware of any other instances in which these additional instructions have been implemented.

Cost variable

Participants were randomly assigned to either a control condition, in which there was no cost for violating Rule 3, or a condition in which violating Rule 3 resulted in a penalty. Rule 3 states that if the orcs outnumber the hobbits on either bank of the river the orcs will then kill the hobbits. In the control condition, if a participant violated Rule 3 he or she was notified via a message box and was then allowed to continue working on the problem. In the cost condition, if a participant violated Rule 3, then the screen turned black, except for some brief instructions, a text box, and a button labeled “Go.” The participant was instructed to click on the “Go” button, and 10 words would then appear on the screen at a rate of 3 s per word. Participants rated the pleasantness of each word on a 5-point scale (1 = *very unpleasant*; 3 = *neutral*; 5 = *very pleasant*). After completing the rating task, participants were returned to the problem state where they committed the violation.

Procedure

Tutorial phase

Participants read a cover story for the hobbits and orcs problem and learned the rules (see Appendix). Before moving on, each participant was required to correctly recite all the rules from memory without error. Once the participant was able to recite the rules, the tutorial phase began.

During the tutorial phase, participants were shown an example problem on the computer, and they were able to practice making moves with the mouse. During the tutorial phase, participants were 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. Next, they were instructed to move one hobbit and one orc to the right bank and then back to the left bank. Next, they received instructions on violating each of the three rules in succession and demonstrated how to do so. The tutorial was complete when participants correctly described all the illegal moves and demonstrated each, as well as making three legal moves and stating the goal of the problem.

After completing the tutorial, pictures of hobbits and orcs were displayed on the computer screen with different combinations of the travelers appearing every 1.2 s for a total of 14.4 s, and participants had to say the travelers aloud. Participants in the silent condition were told that this was to help them better understand the characters so that they could follow the rules more easily. However, the main purpose of the task was to get those in the think aloud condition to talk more often while working on the task and to refer to the characters by their proper names so that the experimenter (Martin E. Knowles) could obtain a better sense of some of the moves that the participants were considering.

Finally, participants were instructed to keep working on the practice problem for 90 s, which consisted of two hobbits and two orcs (instead of three of each as in the experimental problem). Participants in the silent condition did so quietly, while those in the think aloud condition talked aloud.

Participants in the cost condition were next notified of the penalty for violating Rule 3. In both 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-min time limit, that participant was then assisted in finishing the problem and his or her data were not included in the analyses. The maximum solution time of 20 min was chosen to restrict the session length to 1 hr. Participants who completed the problem before time expired completed the operation span (OSPAN) task assessment of working memory (Kane, Bleckley, Conway, & Engle, 2001; Turner & Engle, 1989).

Illegal Moves Made and Considered

The focus of this experiment was on the illegal moves committed by participants. An illegal move was a violation of the third rule, which allowed the hobbits to be eaten by the orcs. Violations of Rule 1 or Rule 2, which involved trying to move the boat while it was empty or attempting to add more than two passengers to the boat, were considered errors reflecting either mistakes in using the interface or a lack of understanding of the rules and thus were not analyzed. Figure 2 shows the hobbits and orcs problem space, including the illegal moves. The problem space is almost completely linear, allowing participants to move either forward or backward at each state. In most states there is only one move that will move the problem solver forward; however, there are two states, State 0 and State 9, where there are two moves that take the problem solver closer to the goal.

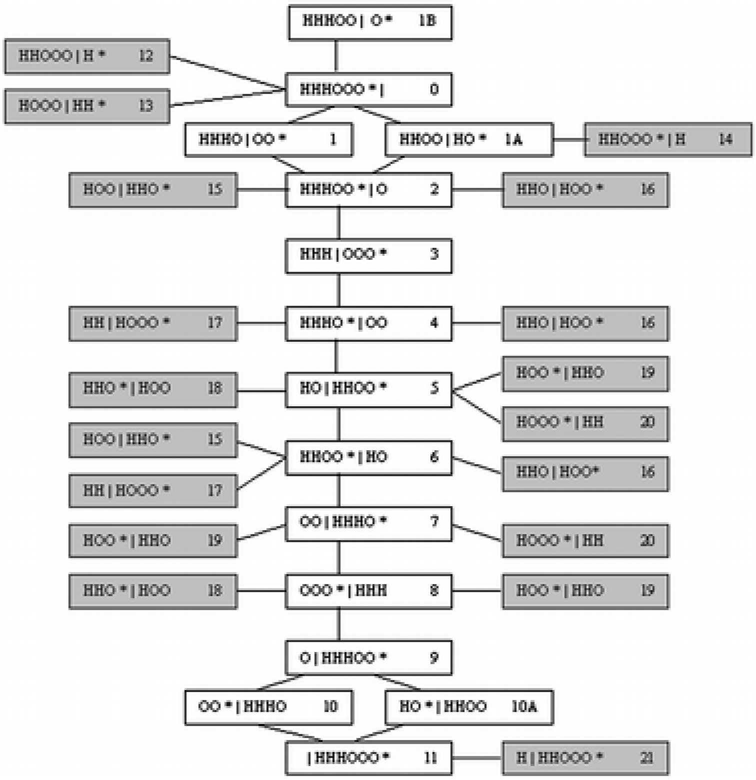

Figure 2. Map of the problem space for the legal and illegal moves of the hobbits (H) and orcs (O) problem. States are labeled inside the right of each rectangle, a white rectangle represents a legal state and a gray rectangle represents an illegal state. State 0 was the start state and State 11 the goal state. Within each state, the boat is represented by an asterisk; the river, a vertical line

Participants in the think aloud condition had Martin E. Knowles attend to their verbal protocols. In the think aloud condition, Knowles tracked the considered-then-made illegal moves and the considered-then-not-made illegal moves and their times in minutes and seconds by attending to the verbal protocol and a timer, marking each occurrence on a score sheet. Knowles coded all of the illegal moves—Rule 3 violations—that the participant considered but did not make.

An illegal move was classified as considered if and only if Martin E. Knowles was able to understand the exact move being considered, the participant realized that the move was illegal, and the participant rejected the move and did not make that move. This could be assessed by watching the participant's moves and attending to his or her verbal protocols. More specifically, considered illegal moves could be observed through removal of travelers from the boat and placement of the mouse over a traveler while the participant stated (a) that making that move was not allowed, (b) that doing so would violate the rules, or (c) any other words that indicated that he or she knew that moving that specific traveler would result in a violation of Rule 3. Broad statements such as "Any move I make will violate Rule 3" or "Anything I do will kill the hobbits" were not counted as considered illegal moves; the statement had to refer to a specific traveler and a specific move.

To verify that the considered illegal move measures were accurate, protocols from 8 different participants were each coded by two different experimenters. Martin E. Knowles coded all eight protocols, and two research assistants coded four other protocols each. A conditional interrater reliability measure based only on illegal moves considered but subsequently rejected was calculated by taking the total number of moves marked as considered by any of the coders and dividing by the number of agreements. We obtained interrater reliabilities greater than 85% for both pairs of coders, suggesting good agreement. For this reason all subsequent protocols were coded by Knowles only.

Results and Discussion

We first consider whether our modified verbal reports affected participants' performance. Next, we analyze participants' performance on specific states to test whether Jeffries et al.'s (1977) findings about what states were most difficult in the problem held true for our interface as well. Third, we consider how individual differences could be understood in terms of our framework. Finally, we present results pertaining to the main focus of the current experiment, namely the effects of the cost manipulation on illegal move rates.

Effects of Verbalization

In Experiment 1, we instructed participants to state all the moves that they were considering in an attempt to obtain more elaborate verbal protocols. This was necessary to discriminate whether participants were simply considering fewer and fewer illegal moves with experience or whether they were getting better at correctly detecting and rejecting illegal moves with experience. Normally, asking participants to verbalize their thoughts has no effect on problem-solving strategies (see Ericsson & Simon, 1993, for a review). However, because participants in our experiments were also asked to state the specific travelers they were considering moving at each state, it was important to ensure that the modified concurrent verbal protocol instructions did not alter participants' problem-solving behavior.

We initially conducted a 2 × 2 analysis of variance (ANOVA)—Verbalization (silent vs. aloud) × 2 Cost (cost vs. control)—with both verbalization and cost as between-subjects factors. Means and standard deviations of relevant dependent measures for the silent and aloud groups are reported in Table 1. (Statistical results related to the cost manipulation are presented in the next section to allow focus on the question of the effects of verbalization.) No differences between the silent and aloud groups emerged for legal moves made, $F(1, 60) = 1.78$, $MSE = 227.28$, or for total time to complete the problem, $F(1, 60) = 1.02$, $MSE = 18.87$. We also found no difference for average time per move, illegal moves made, or proportion of illegal moves made (all *F*s < 1). The interactions between cost and verbalization for all dependent measures also proved nonsignificant (all *F*s < 1).

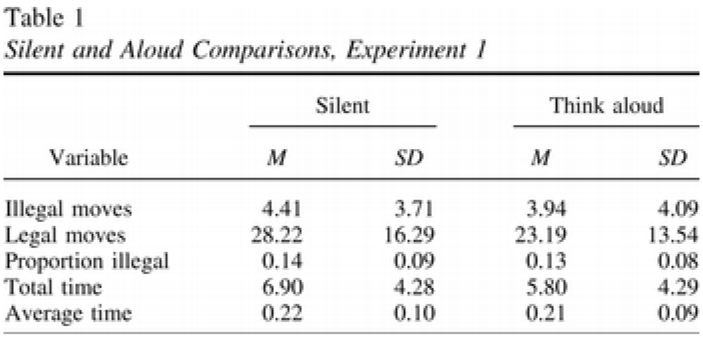

Table 1
Silent and Aloud Comparisons, Experiment 1

Variable	Silent *M*	Silent *SD*	Think aloud *M*	Think aloud *SD*
Illegal moves	4.41	3.71	3.94	4.09
Legal moves	28.22	16.29	23.19	13.54
Proportion illegal	0.14	0.09	0.13	0.08
Total time	6.90	4.28	5.80	4.29
Average time	0.22	0.10	0.21	0.09

Silent and Aloud Comparisons, Experiment 1

In summary, we were unable to detect any differences between the silent and aloud groups on any of our dependent measures. Thus, in subsequent analyses we collapsed over the silent and aloud groups wherever possible. The results suggest that our novel approach to collecting verbal reports did not unduly influence participants' problem-solving behavior. Our approach of asking participants to verbalize more about specific states under consideration may prove useful in other problem-solving applications.

State Versus State Comparisons

The hobbits and orcs problem, although almost completely linear in its solution, has some states that are visited more frequently and thus are presumably harder (Greeno, 1974; Jeffries et al., 1977). In particular, States 2 and 5 generated substantially more visits than the other states. Because our graphical interface was somewhat different than the textual interface used in earlier research, we wanted to verify that the same pattern would emerge. Because previous research has shown that States 2 and 5 are generally the hardest (Greeno, 1974; Jeffries et al., 1977), we conducted paired-samples *t* tests comparing States 2 and 5 with the mean of all the other states in the problem combined for both legal and illegal moves. Means and *t* values for the control group are provided in Table 2 for legal and illegal moves. All *p* values were Bonferroni corrected to control for familywise error.

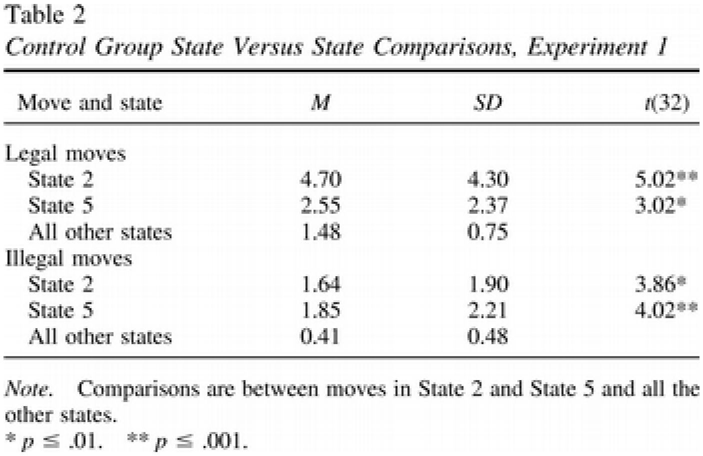

Table 2
Control Group State Versus State Comparisons, Experiment 1

Move and state	*M*	*SD*	*t*(32)
Legal moves			
State 2	4.70	4.30	5.02**
State 5	2.55	2.37	3.02*
All other states	1.48	0.75	
Illegal moves			
State 2	1.64	1.90	3.86*
State 5	1.85	2.21	4.02**
All other states	0.41	0.48	

Note. Comparisons are between moves in State 2 and State 5 and all the other states.
* $p \leq .01$. ** $p \leq .001$.

Control Group State Versus State Comparisons, Experiment 1

Our results largely replicated those of Jeffries et al. (1977) and Greeno (1974) and Thomas (1974). There were significantly more legal visits to States 2 and 5 when their mean legal moves were compared with the mean of the other states. As with legal moves, on average there were significantly more illegal moves created in States 2 and 5 when they were compared with the mean of the other problem states. There were no illegal moves committed in States 1B, 1, 3, 8, 9, 10, 10A, or 11, and therefore they were not included in the mean of illegal moves.

One might wonder why States 2 and 5 generate more visits and errors than the other states. State 2 has two possible illegal alternatives, and it is one of the few states with two moves that backtrack away from the goal. In addition, if a participant does not do at least some minimal planning, then the correct move of two orcs to the right bank, which would make three orcs on the right and no hobbits, may seem like a dead end. Without planning, participants may not realize that one orc would return to the left bank, and if there are three orcs on the right bank and the boat can only hold two travelers, then any move of the hobbits to the right bank would end with the hobbits being outnumbered on the right bank. In State 5, there are three illegal alternatives, and it is the only move in the problem, as noted by Jeffries et al. (1977), in which two travelers must be returned to the left bank. This move of one hobbit and one orc to the left bank seems counterintuitive and more like backtracking than advancing (e.g., Thomas, 1974).

That individual states differ in their likelihood of generating illegal moves creates difficulties for knowing whether participants reduce the number of illegal moves they make as they work on the problem. Even if one was to observe such a pattern, it could be due to a greater number of visits to the most difficult states in the early part of the problem-solving episode rather than actual learning. This issue was addressed more effectively in Experiment 2.

Individual Differences

Individuals clearly differed in the number of illegal moves they made among the control group. (In the cost group, illegal moves were uniformly quite rare.) It makes sense to ask whether our framework can explain individual differences in illegal move rates in the control group. According to the framework, effects at each level must be controlled either experimentally or statistically before the subsequent levels can be addressed. We therefore began by considering the generation-rate hypothesis as an explanation of individual differences. Specifically, the generation-rate hypothesis would predict that legal moves should predict how many illegal moves will be made. The caution and rule-verification hypotheses would both predict that once generation rate is statistically controlled, correct rejections of illegal moves should be inversely related to the number of illegal moves committed.

We performed a regression analysis for the control group, with illegal moves as the dependent factor and legal and considered moves as the independent factors. The number of legal moves was entered first and the number of considered moves was entered second. Legal moves accounted for significant variance, $R^2 = .545$, $F(1, 14) = 16.75$, $p < .01$, supporting the generation-rate hypothesis. Specifically, as more opportunities to make illegal moves (in terms of moves) became available, more illegal moves were actually made. Once generation rate was controlled for, the number of considered but correctly rejected moves was inversely related to illegal moves made, $\Delta R^2 = .244$, $F(1, 13) = 15.01$, $p < .01$. The regression equation was $I = .462L - .652C - 1.97$, where I is the number of illegal moves, L is the number of legal moves, and C is the number of correctly rejected considered illegal moves. The regression analysis therefore supports the idea that generation rate alone is not responsible for the individual differences observed in the control group, though generation rate accounted for a large proportion of the variance.

Another individual-differences question concerns the relationship between working memory capacity as assessed with OSPAN and various solution characteristics. We summarize these results only briefly by saying that we found no significant correlations between OSPAN and any of our solution characteristics in Experiment 1. More details on working memory and our puzzles are given following Experiment 3 in a separate section (see Working Memory Assessment).

Effects of the Cost Manipulation on Illegal Moves

The cost manipulation either penalized participants (cost) or allowed them to continue working on the problem (control) after making an illegal move. Our most important question in Experiment 1 was whether this manipulation affected participants' legal and illegal move making. If participants' solutions could be improved without manipulating the problem itself or the resources required to solve the problem, the cost manipulation could prove useful for improving problem-solving performance.

Means and standard deviations of relevant dependent measures for the cost and control groups are reported in Table 3. Independent-samples *t* tests revealed neither significant differences for number of illegal moves considered (but not made) nor for legal moves made (both *t*s < 1). We were also unable to detect any differences in total solution time across the two groups, *t*(61) = 1.58. However, significant differences emerged in the number of illegal moves committed, *t*(61) = 3.05, *p* <.01, indicating that those in the cost condition made fewer illegal moves compared with those in the control condition.

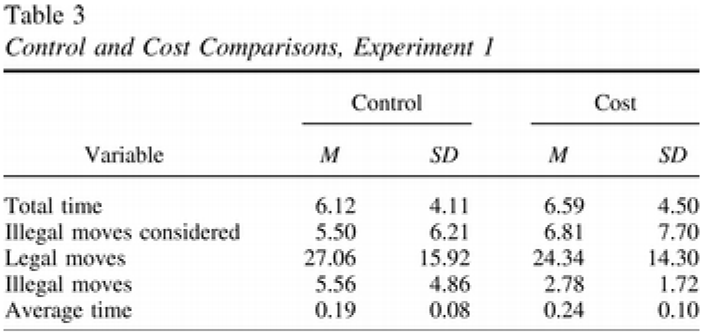

Table 3

Control and Cost Comparisons, Experiment 1

	Control		Cost	
Variable	*M*	*SD*	*M*	*SD*
Total time	6.12	4.11	6.59	4.50
Illegal moves considered	5.50	6.21	6.81	7.70
Legal moves	27.06	15.92	24.34	14.30
Illegal moves	5.56	4.86	2.78	1.72
Average time	0.19	0.08	0.24	0.10

Control and Cost Comparisons, Experiment 1

The most important result of these analyses was that the cost manipulation reduced illegal moves without affecting legal moves. Because illegal moves were reduced but legal moves were not, it seems unlikely that the cost manipulation worked merely by reducing the number of decisions that participants made. Thus, the generation-rate hypothesis cannot account for the positive effects of the cost manipulation, suggesting that rule verification or caution is the source of its effects on illegal moves.

Experiment 2

Experiment 1 demonstrated that the cost manipulation reduced the number of illegal moves that participants generated. In Experiment 2, we addressed whether the cost manipulation would produce sustained benefits when participants solved the problem a second time. After they solved the problem once, we canceled the cost instruction by informing participants that there would be no more penalty for making illegal moves.

A secondary goal was to learn more about the causes of improvement in illegal move rates with practice while avoiding confounds relating to some states being more difficult than others. Our control manipulation (in which no costs were imposed) closely replicated an earlier study by Reed et al. (1974) in which participants solved the same problem twice. By solving the same problem twice, the same states would be encountered during both solutions. This allowed us to compare the relative change in legal and illegal moves with practice on the same problem states.

Method

Participants

Participants were 91 undergraduates from the University of Florida who received course credit for their participation. Ten participants proved unable to solve either problem in 20 min. They were assisted in completion of the problem, and thus they were not included in the analysis. The analysis included the remaining 81 participants.

Problem

The problem was the same as that used in Experiment 1.

Design

The experiment had a 2 × 2 design: Cost (cost vs. control) × Solution Attempt (first vs. second), where cost was a between-subjects factor and solution attempt was a within-subjects factor. There was no verbalization factor; all participants solved the problems silently.

Cost variable

As in Experiment 1, participants in the cost group were told that there would be a penalty for violating the rules in which they would have to rate 10 words for pleasantness every time they made an illegal move. However, on the second solution attempt, participants were informed that there would be no more penalty for making illegal moves. The control group never received any instructions about avoiding illegal moves and suffered no penalty for making illegal moves.

Solution attempt variable

Each participant solved the same hobbits and orcs problem twice. Each problem had to be solved within the 20-min time limit, and the second solving of the problem was always done under the control rules (i.e., no penalty) for making an illegal move.

Procedure

The procedures followed those of Experiment 1, except that after the tutorial phase participants did not have to view combinations of the travelers on the screen. Also, after solving the problem the first time the participant was then instructed to solve the same hobbits and orcs problem a second time. After solving the problem a second time, the participant was asked to complete the OSPAN task. The cost group solved the first problem under the cost directions but solved the second problem under the control directions. Those in the control group solved both problems under the control directions.

Results and Discussion

In Experiment 2, participants solved the hobbits and orcs problem twice under either the cost or control manipulation. We were interested to see whether we could replicate the findings of Experiment 1: the cost manipulation decreasing illegal but not legal moves. We also wanted to see whether participants would show decreases in illegal moves as they gained experience with the problem. We wondered whether participants would continue to produce fewer illegal moves even after the penalty for making them was removed on the second solution attempt or whether they would revert to higher rates of illegal move making (see also Reed et al., 1974).

Analysis of Legal Moves

We conducted a 2 × 2 mixed factorial ANOVA—Cost (cost vs. control) × Solution Attempt (first vs. second problem)—on the legal moves for Experiment 2. Cost was a between-subjects factor, and solution attempt was a within-subjects factor. The analysis of legal moves revealed no main effect of cost, $F(1, 79) = 1.90$, $MSE = 346.15$; no main effect of solution attempt, $F(1, 79) = 2.08$, $MSE = 223.97$; and no interaction between cost and solution ($F < 1$).

These findings replicated those of Experiment 1 in that legal moves did not differ between the cost and control groups. In addition, legal moves did not decrease from the first solution to the second. If illegal move rates were to be affected by the cost manipulation but legal moves were intact, our assertion that the changes in illegal move rates due to the cost manipulation could not be attributed to the generation-rate hypothesis would be supported. A similar conclusion would be possible for practice effects—because legal move rates were unchanged by repeated solution, changes in illegal moves would be difficult to attribute to differences in the number of opportunities to produce illegal moves.

Analysis of Illegal Moves

We also conducted a 2 × 2 mixed factorial ANOVA—Cost (cost vs. control) × Solution Attempt (first vs. second problem) for illegal moves. Figure 3 shows the number of illegal moves as a function of cost and solution. We detected a main effect of cost, $F(1, 79) = 10.30$, $MSE = 66.81$, $p < .01$, indicating that the cost group ($M = 2.35$, $SD = 2.01$) made fewer illegal moves than the control group ($M = 6.48$, $SD = 7.97$). We also detected a main effect for solution attempt, $F(1, 79) = 4.58$, $MSE = 28.89$, $p < .05$, indicating that participants made fewer illegal moves on the second ($M = 3.51$, $SD = 4.26$) compared with the first problem ($M = 5.32$, $SD = 9.28$). Although the two-way interaction was not significant, it approached significance, $F(1, 79) = 3.12$, $p = .08$. Whereas the control group apparently produced fewer illegal moves on the second solution attempt than on the first, $t(39) = 2.00$, $p = .05$, there were no changes in illegal moves for the cost group ($t < 1$). This result may reflect a floor effect such that participants had little room for improvement in the cost condition.

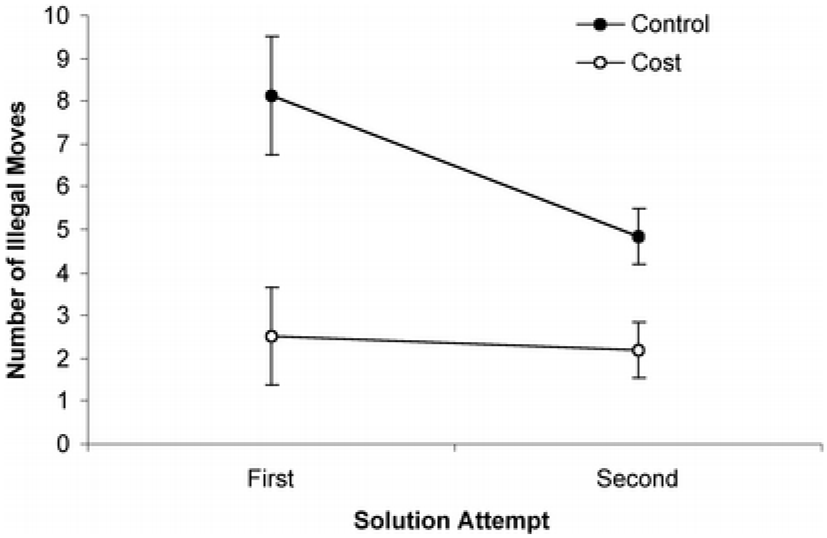

Figure 3. Number of illegal moves for Experiment 2 by condition and solution attempt. Error bars represent ± SE

In sum, the number of legal moves was not affected by the cost manipulation (and the number of legal moves also did not change as participants gained experience with the problem). However, illegal moves decreased with practice, replicating an earlier result by Reed et al. (1974). Illegal moves also decreased following the cost manipulation. Even though the penalty for illegal move making was lifted in the cost group during the second solution attempt, the rate of illegal move making remained low. This suggests that caution or rule verification or both was likely the cause of the benefits of both the cost manipulation and the practice effect.

Experiment 3

In Experiment 3, we replaced the second solution of the hobbits and orcs problem with an isomorphic problem. An *isomorph* is a problem that has the same underlying structure or problem space as another problem, although it may not have any outwardly obvious similarities. Whereas our earlier experiments showed that the generation-rate hypothesis often provided a good explanation for factors such as individual differences in illegal move making or differences in the difficulty of particular states, it could not easily account for the benefits of the cost manipulation on illegal move generation. Therefore, the transfer effects observed in Experiment 2 may have demonstrated increasing caution (the caution hypothesis) or increasing efficiency at checking the rules (the rule-verification hypothesis). By changing the problem between the two solutions, rule-specific transfer would presumably be eliminated. Finding transfer following the cost manipulation would therefore tend to favor the caution hypothesis because participants could not be transferring knowledge of how to check particular rules across two very different problems. In earlier research, Reed et al. (1974) found no transfer between dissimilar isomorphs, suggesting that caution was not the likely explanation for standard practice effects. Therefore, it would be intriguing to find a different pattern of results following the cost manipulation.

Method

Participants

Participants were 86 undergraduates from the University of Florida who received course credit for their participation. Nine participants proved unable to solve one or both problems in 20 min. They were assisted in completing the problem and thus were excluded from analysis. The remaining 77 participants were submitted to analysis.

Problems

Hobbits and orcs

The hobbits and orcs problem was the same as that used in Experiments 1 and 2.

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 could 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 and is depicted in Figure 4. A red “Remove” or “Add” button appeared on the screen indicating to participants whether they were to be removing or adding isotopes at each step. When the “Remove” button was presented participants 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 two large blue and orange beakers. 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.

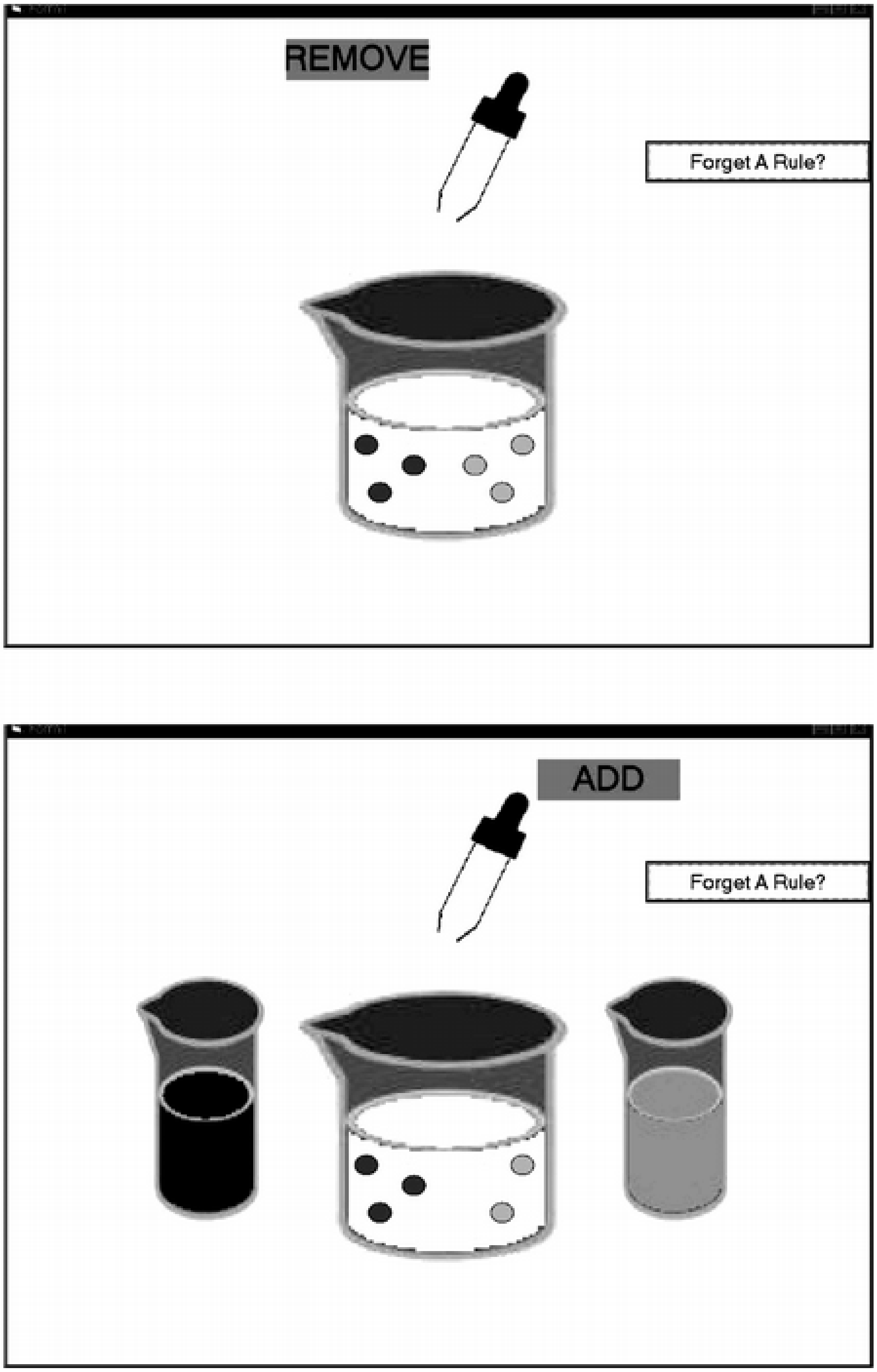

Figure 4. Display seen by participants for the titration problem. The top panel is the initial state of the problem, where participants were to click on isotopes in the white beaker to make the isotopes appear in the dropper. Clicking on the "Remove" button would clear the dropper and the dark (blue) and light (orange) colored beakers would appear. The bottom panel would appear after removing one light colored isotope. Participants would then click on the light and dark colored beakers to deposit isotopes into the dropper. After doing this they would click on the "Add" button, and the isotopes in the dropper would appear in the white beaker and the light and dark colored beakers would disappear

The rules stated that it was impossible to have more than three blue or orange isotopes on the screen at any time. If participants 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 they were notified via a message box, and the move did not occur.

Results and Discussion

For all subsequent analyses, we conducted a 2 × 2 × 2 mixed factorial ANOVA: Cost (cost vs. control) × Order (titration first vs. hobbits first) × Solution (first vs. second problem). Cost and order were between-subjects factors, and solution was a within-subjects factor. Means and standard deviations for legal moves can be found in Figure 5, and Figure 6 contains those for illegal moves.

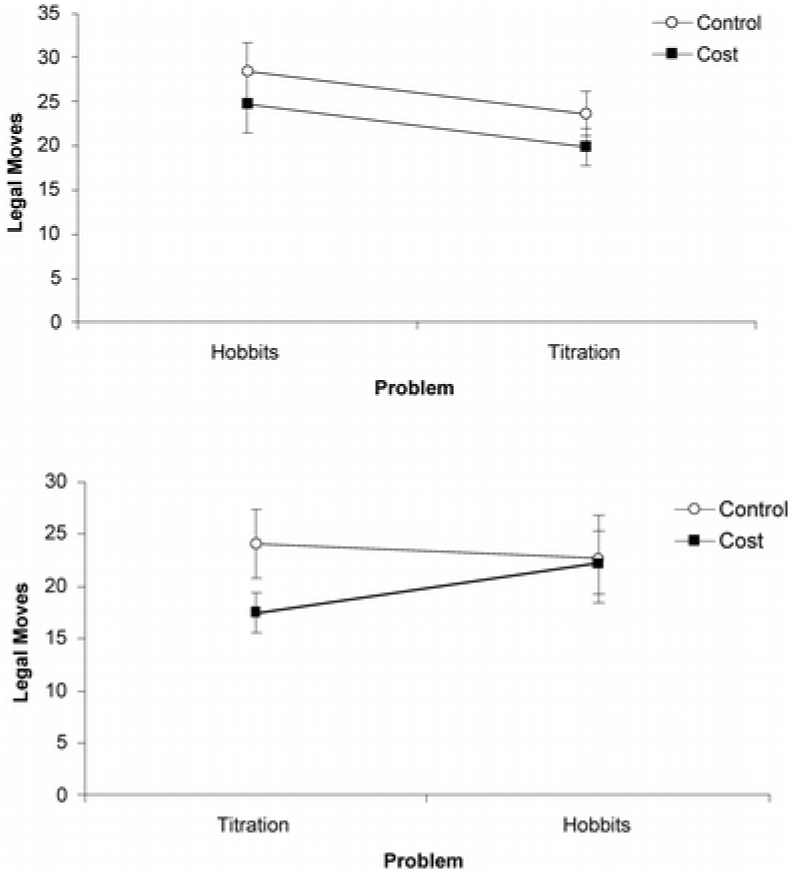

Figure 5. Number of legal moves for Experiment 3 for both the control and cost groups. The top panel represents those participants who solved the hobbits and orcs problem first, followed by the titration problem. The bottom panel represents those who solved the titration problem first, followed by the hobbits and orcs problem. Error bars represent ± SE

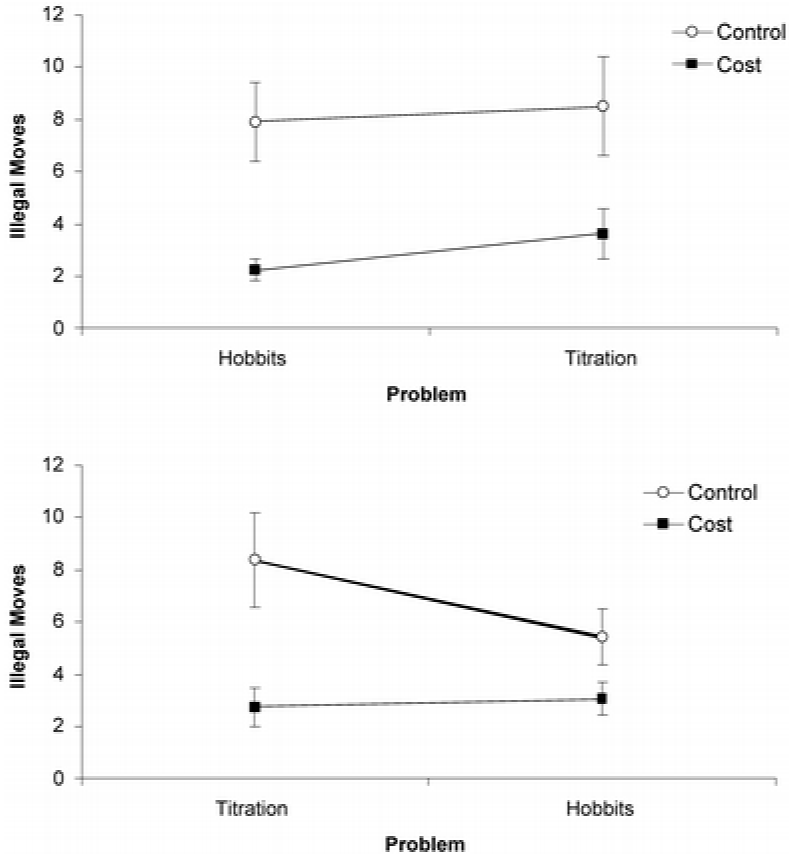

Figure 6. Number of illegal moves for Experiment 3 for both the control and cost groups. The top panel represents those participants who solved the hobbits and orcs problem first, followed by the titration problem. The bottom panel represents those who solved the titration problem first, followed by the hobbits and orcs problem. Error bars represent ± SE

Analysis of Legal Moves

The analysis of legal moves failed to detect a main effect of cost, $F(1, 73) = 2.56$, $MSE = 191.92$; a main effect of order, $F(1, 73) = 1.32$, $MSE = 191.92$; or a main effect of solution ($F < 1$). No interactions emerged between cost and solution ($F < 1$) or between order and solution, $F(1, 73) = 2.38$, and there was no three-way interaction ($F < 1$).

The lack of a main effect of order indicates that participants' performance on both problems was roughly equivalent on the first solution attempt. The lack of a main effect of cost seems to support the finding from the first two experiments that the cost manipulation did not have much effect on participants' legal moves. The lack of a main effect of solution suggests little cross-problem transfer in terms of legal moves (see also Reed et al., 1974). To further assess differences between the two isomorphs, we calculated the legal moves made per state by dividing the number of legal moves made by each participant by the number of legal problem states. An independent-samples *t* test revealed that the mean number of legal moves participants made per state did not differ between the two isomorphs, $t(38) = 0.71$, $p > .05$, indicating that legal moves occurred approximately equally often in both problems.

Analysis of Illegal Moves

An illegal move was a violation of the third rule, which created unstable isotopes by having an unequal number of blue and orange isotopes in the white beaker when the number of blue isotopes did not equal zero or three.

As in Experiments 1 and 2, violations of Rule 1 or Rule 2, which involved trying to add or remove isotopes when the dropper had either zero or more than two isotopes, were considered errors and were not analyzed.

For illegal moves, we detected a main effect of cost, $F(1, 73) = 17.70$, $MSE = 45.95$, $p < .01$, indicating that the cost group ($M = 2.92$, $SD = 2.36$) made fewer illegal moves than the control group ($M = 7.58$, $SD = 6.19$). However, the main effects of order and solution were both nonsignificant (*F*s < 1). The lack of a main effect of order indicated that performance was roughly equivalent on the first solution attempt. No interactions emerged between cost and solution, $F(1, 73) = 2.26$, or between order and solution, $F(1, 73) = 2.87$, and there was no three-way interaction ($F < 1$). To further assess differences between the two isomorphs, we calculated the illegal moves made per state by dividing the number of illegal moves made by each participant by the number of legal problem states. An independent-samples *t* test revealed that the mean number of illegal moves participants made per state did not differ between the two isomorphs, $t(38) = 0.20$, $p > .05$.

We note that the benefits of the cost manipulation persisted even in the absence of a penalty on the second problem, indicating that the benefits transferred to a novel problem-solving task. For the cost group, a paired-samples *t* test revealed no significant difference between the number of illegal moves on the first problem compared with the second, $t(36) = 1.38$, $p > .05$, indicating that participants continued to make a minimal number of illegal moves even once they no longer received the cost penalty for making illegal moves. An independent-samples *t* test comparing the cost versus control groups on their second solution revealed significant differences in the number of illegal moves on the second problem, $t(75) = 2.78$, $p < .01$, indicating that the cost group ($M = 3.38$, $SD = 3.68$) was still making fewer illegal moves than the control group ($M = 7.02$, $SD = 7.14$) even after the cost was removed. This suggests that the cost manipulation may have its effects by increasing general cautiousness in some way.

As in Experiments 1 and 2, the cost manipulation reduced the number of illegal moves without changing the number of legal moves. Participants also continued to avoid illegal moves in the cost condition even after the penalty was removed. Of interest, the magnitude of the sustained benefits resulting from the cost manipulation was approximately the same size for a novel problem ($\eta^2 = .094$) as it was for the same problem being solved a second time ($\eta^2 = .096$), suggesting that the transfer effect was mediated by caution or some other factor that was not problem-specific or practice at verifying particular rules. This does not necessarily imply there is no contribution of rule learning to the effects of the cost manipulation. However, it appears that the sustained benefits of the cost manipulation are primarily caused by something that can be readily transferred to novel problems.

In contrast, we found that practice on one problem did not reduce legal nor illegal move making on the second problem, replicating Reed et al.'s (1974) results. However, practice on the same problem twice did improve illegal move rates (see Experiment 2). This suggests that the practice effect was not likely due to nonspecific changes in cautiousness but rather to enhanced rule-verification skills. Given that this latter conclusion is based on a null result, caution is warranted in accepting this claim. Additional research would be necessary to understand the source of the practice effect more fully. To summarize, then, we replicated Reed et al.'s results in a control group, but our cost manipulation showed an alternative pattern of results that suggested transfer even across isomorphs.

Working Memory Assessment

It has been argued that resource limitations are a major factor contributing to a person's selection of illegal moves because resources constrain the ability to correctly calculate future states or even remember to check a move for legality (Jeffries et al., 1977). We believe that working memory is one of the resources that Jeffries et al. (1977) may have been referring to because working memory is thought to be a central resource in problem solving and therefore should be implicated when people calculate future states and check moves for legality (Kotovsky & Fallside, 1989).

In all of the experiments, after completing the final target problem, the OSPAN task was administered to assess working memory span. To increase the power to detect a correlation between working memory capacity and illegal move commission, we also included data from another experiment that was identical to Experiment 1, except there was no cost manipulation—all participants completed the problem under control instructions. Further information on this study, whose methods and results closely replicated the control condition of Experiment 1, can be found in Knowles (2004). A total of 300 participants successfully completed the target problem(s) in the four experiments. However, 18 participants in the cost group and 14 in the control group did not have enough time at the end of their experiment to complete the OSPAN task. This resulted in OSPAN scores for 172 participants in the control group and 96 participants in the cost group.

We conducted correlation analyses on the cost and control groups to compare OSPAN scores and the number of legal and illegal moves committed on each participant's first solution. The analysis revealed that legal moves were not significantly correlated with OSPAN for either the control (r =.025, p =.746) or cost (r =.013, p =.903) conditions. We also did not detect a reliable correlation between illegal moves and OSPAN in the control (r = −.132, p =.083) or cost (r = −.167, p =.103) conditions. This suggests that the resource limitation hypothesis may not provide the only explanation for errors in river-crossing problems such as the hobbits and orcs problem. Instead, we propose that other factors (e.g., caution) may influence the selection of illegal moves, at least to some degree. Although working memory capacity likely has a much greater impact on other kinds of problem solving (e.g., Kotovsky et al., 1985; Phillips, Wynn, Gilhooly, Della Sala, & Logie, 1999; Welsh, Satterlee-Cartmell, & Stine, 1999), it does not appear to drive the effects in this experiment based on these results.

General Discussion

Illegal moves are only rarely studied despite the fact that they often account for a significant proportion of the total moves that people make. On the hobbits and orcs problem, for example, illegal moves constituted 21.5% of the total moves that people made in the control group and 10.0% in the cost group. This was similar to the findings of Jeffries et al. (1977), in which percentages ranged from 14.75% to 32.8% in the various isomorphs. Therefore, methods that reduce illegal moves can shorten solutions and result in more efficient problem solving. The impact can be even greater when participants are required to plan and are not allowed to actually make the moves and observe the consequences. When people plan ahead, the negative impact of illegal moves is even greater because any move considered after an illegal move produces wasted effort (Delaney et al., 2004).

Effects of the Cost Manipulation

Relatively simple manipulations can affect the way participants approach a problem. For example, introducing a cost for making each move increases planfulness (O'Hara & Payne, 1998). Our main discovery was that a simple cost manipulation applied to illegal moves produced reductions in illegal move rates that remained when the cost was removed (Experiments 1–3) and that transferred even to novel problems (Experiment 3). It appears that this was primarily due to an increase in the probability of checking the legality of moves. It remains to be seen how durable the changes in cautiousness are, as we tested only immediate transfer. It would be interesting to see whether the cost manipulation produces changes that last a week or more, as other problem-solving manipulations do (e.g., preplanning practice; Delaney et al., 2004).

We also found that individuals in the control group (who received no cost manipulation) differed in their likelihood of rejecting considered illegal moves. In Experiment 1, the number of moves that were correctly rejected was inversely related to the number of illegal moves actually committed after factoring out generation rate. This could have been due to individual differences in efficiency at checking the rules or to differences in caution. Davies (2003) showed that there are stable individual differences in the amount of initial versus concurrent planning that people do on various problems. We might speculate that cautiousness is also an individual difference in problem solving. It is worth discovering in future work whether cautiousness is a stable individual difference. If so, then our cost manipulation may have the effect of converting incautious individuals into cautious individuals.

Jeffries et al. (1977, p. 437) assumed that the majority of the differences in illegal move rates between isomorphs had to do with available cognitive resources or knowledge of the rules. The cost manipulation significantly manipulated checking rates and carried the majority of the variance in our river-crossing problems. This finding could indicate that cognitive resources play a relatively minor role in such problems because if they did not then the cost manipulation should have produced much smaller effects than it did. We were also unable to detect evidence that mental resources—as assessed by standard working memory measures—could explain much of the individual differences in illegal move making on the hobbits and orcs task. In other problems in which the central difficulty is maintaining multiple alternatives in memory or multiple goals, resources would likely be much easier to detect (for a review, see Phillips & Forshaw, 1998). However, because approximately 10% of the participants did not complete the problem within the time allotted, cognitive resources may play a somewhat larger role in problem-solving performance than these results indicate. It may be that nonsolvers were among the lowest span individuals and that including them could have increased the correlations with working memory capacity.

Application of the Framework to Previous Research

Using the process model of Jeffries et al. (1977) as a basis, we derived a generalized framework for examining the source of illegal moves during problem solving. Our suggestion is that this framework can be used to understand why a manipulation that reduces illegal moves produces its effects. Manipulations that primarily either reduce the number of total moves (and hence the number of opportunities to produce an illegal move) or that change strategies such that illegal moves are considered less frequently in any given state would have their impact in Stage I of the framework. Manipulations that affect cautiousness (i.e., the rate of remembering to check the rules) would have their impact on Stage II, as our cost manipulation did. Finally, manipulations that affect the ease of verifying the rules would have their impact in Stage III. The impact of a manipulation that affects a particular stage is dependent on the likelihood of reaching that stage. In cases in which illegal moves never come to mind, cautiousness would have no impact, and in cases in which people never remember to check the rules the difficulty of verifying the rule would similarly have no impact.

We can analyze earlier research findings in terms of this framework as well. For example, although checking the rules in hobbits and orcs is relatively simple compared with some other problems, changes in the ease of checking can produce changes in illegal move rates in some cases. Jeffries et al. (1977), for example, argued that one of their isomorphs ("elves and men II") made it easy to check the rules because all rule violations were apparent on both banks of the river and that illegal move rates were reduced as a consequence. However, we believe that altering the rules in this fashion operates on Stage II of the framework by eliminating the need to be cautious. Because the tests can all be made perceptually, this is not cautiousness in the sense that we mean it (i.e., remembering to do the check of the rules), rather it is a way of eliminating the need for cautiousness by making the rule violations perceptually obvious.

Making rules external eliminates the possibility of illegal moves involving those rules (Zhang & Norman, 1994). Making rules external in our model really eliminates the need to be cautious by making the errors perceptually available. That is why error rates drop to zero or near zero in those cases. Only when people need to remember to check the rules (i.e., be cautious) do illegal moves occur. In the same fashion, presenting all legal options stops the Stage I generation of any illegal moves, eliminating the need to be cautious or check the rules and therefore reduces illegal moves to near zero (Kotovsky & Simon, 1990).

The control conditions of our Experiments 2 and 3 were very close replications of earlier experiments by Reed et al. (1974). In Experiment 2, we found that solving the same hobbits and orcs problem twice reduced illegal moves but had little effect on legal moves, replicating the results of their Experiment 2. We also found that transfer between isomorphs was very limited, despite practice effects on the same isomorph, replicating their Experiment 1. Taken together, these results suggest that caution cannot easily explain the gains observed following repeated solving of the same problem. We therefore concur with Reed et al.'s suggestion that more effective legality checking (i.e., rule verification) is likely responsible for the transfer effects on repeated solutions of the same problem.

Conclusions

We found that encouraging people to verbalize more information about the moves they were considering did not affect problem-solving performance, which may prove useful in future studies. However, our most important finding was that the cost manipulation not only reduces illegal move making but continues to do so even on a novel problem, suggesting general transfer. This finding seems to indicate that problem solvers may be relying on inefficient problem-solving strategies when they encounter problems for the first time, but with minor manipulations, efficient problem-solving performance can be increased relatively easily. These results also indicate that some general learning, which may be cautiousness, can be obtained, sustained, and transferred to a novel problem, at least for a brief period of time. Some limitations of this work are that we were unable to assess how long these benefits can be sustained and the extent to which they can be transferred to novel problems. We also replicated earlier results by Reed et al. (1974) showing that practice does not show the same pattern of benefits on novel problems, suggesting rule-specific transfer. Correlations with working memory capacity, despite relatively large numbers of participants, showed minimal relationships with either legal or illegal move making. However, because not all participants were able to solve the problem we must view this last finding with some reservation and look to future research to help address these issues.

In sum, it appears that with minimal adjustment participants can learn to avoid illegal moves and become more efficient problem solvers. Increasing the cost of illegal move making, even slightly, changes behavior and improves problem-solving performance. Whenever one plans ahead, one needs to be careful not to generate options that are impossible given the situation. Impossible (or illegal) moves may be generated whenever externally imposed rules, such as traffic laws, office procedures, or customs, prohibit the application of a seemingly promising action. People may violate such rules because they fail to check for them or because they fail to properly envision the consequences of the action. Puzzles like the hobbits and orcs problem create a laboratory analogue of this everyday phenomenon whereby practical ways of reducing rule violations can be discovered.

References

Atwood, M. E., Masson, M. E. J., & Polson, P. G. (1980). Further explorations with a process model for water jug problems. *Memory & Cognition*, *8*, 182–192.

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

Davies, S. P. (2000). Move evaluation as a predictor and moderator of success in solutions to well-structured problems. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *53*(A), 1186–1201.

Davies, S. P. (2003). Initial and concurrent planning in solutions to well-structured problems. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *56*(A), 1147–1164.

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.

Greeno, J. G. (1974). Hobbits and orcs: Acquisition of a sequential concept. *Cognitive Psychology*, *6*, 270–292.

Griggs, R. A., & Cox, J. R. (1983). The effects of problem content and negation on Wason's selection task. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *35*(A), 519–533.

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. (2004). *Illegal moves as a source of problem difficulty*.Unpublished master's thesis, University of Florida, Gainesville.

Kotovsky, K., & Fallside, D. F. (1989). Representation and transfer in problem solving. In D.Klahr & K.Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 69–108). Hillsdale, NJ: Erlbaum.

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.

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

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.

Phillips, L. H., & Forshaw, M. J. (1998). The role of working memory in age differences in reasoning. In R. H.Logie & K. J.Gilhooly (Eds.), *Working memory and thinking* (pp. 23–43). Hove, England: Psychology Press.

Phillips, L. H., Wynn, V., Gilhooly, K. J., Della Sala, S., & Logie, R. H. (1999). The role of memory in the Tower of London task. *Memory*, *7*, 209–231.

Reed, S. K., Ernst, G. W., & Banerji, R. (1974). The role of analogy in transfer between similar problem states. *Cognitive Psychology*, *6*, 436–450.

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

Saariluoma, P. (1992). Error in chess: The apperception-restructuring view. *Psychological Research*, *54*, 17–26.

Thomas, J. C. (1974). An analysis of behavior in the hobbits-orcs problem. *Cognitive Psychology*, *6*, 257–269.

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

Welsh, M. C., Satterlee-Cartmell, T., & Stine, M. (1999). Towers of Hanoi and London: Contribution of working memory and inhibition to performance. *Brain and Cognition*, *41*, 231–242.

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

APPENDIX

APPENDIX A: Cover Stories

Hobbits and Orcs: The problem consists of a river, a boat, and 6 travelers. The travelers are 3 Hobbits and 3 Orcs; your job is to get all 6 travelers to the right side of the river safely. However, there are 3 rules that you must follow. Rule 1: There must be at least 1 traveler in the boat to cross the river in either direction. Rule 2: The boat can only hold a maximum of 2 travelers at one time. Rule 3: At no time may the Orcs outnumber the Hobbits on either bank of the river or the Orcs will then kill the Hobbits. Before moving on take some time to learn the rules so that you can say them twice in a row without error. Let the experimenter know when you are ready.

Titration: The problem consists of a beaker, a dropper, and 6 isotopes. The isotopes are 3 Blu-ites and 3 Orange-ites; your job is to remove all 6 isotopes from the beaker without having them explode. However, there are 3 rules that you must follow. Rule 1: There must be at least 1 isotope in the dropper to remove or add anything from the beaker. Rule 2: The dropper can only hold a maximum of 2 isotopes at one time. Rule 3: If the number of Blu-ites in the beaker does not equal either the maximum (3) or the minimum (0) number possible then there must be an equal number of Blu-ites and Orange-ites in the beaker at all times. Before moving on take some time to learn the rules so that you can say them twice in a row without error. Let the experimenter know when you are ready.