LEROUX, ANDREW L., M.A. Problem Solving Difficulty From Unrecognized Equivalency (2011)
Directed by Dr. Peter F. Delaney. 85 pp.

Equivalency occurs in problem solving when a single state can be represented by multiple configurations. If unrecognized, multiple equivalent states can be perceived as new, unique states. This confusion can hinder problem solving. Three experiments used two river crossing problems to explain changes in difficulty through changes in the external representation. The order in which objects were selected, the appearance of objects, and the spatial location of objects were manipulated to affect the difficulty of recognizing equivalent problem states. Blocking equivalency by fixing the selection order was the most reliable effect, helping participants solve the problems in fewer moves.

PROBLEM SOLVING DIFFICULTY FROM UNRECOGNIZED EQUIVALENCY

by

Andrew L. LeRoux

A Thesis Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Greensboro
2011

Approved by

____

Committee Chair

APPROVAL PAGE

This thesis has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair ______________________
Peter F. Delaney

Committee Members ______________________
Dayna R. Touron

____
Edward J. Wisniewski

3/22/2011
Date of Acceptance by Committee

2/12/2010
Date of Final Oral Examination

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Solving algebra problems is challenging for novices. While searching for a solution, a student may change an expression from "$xy = z$" to "$yx = z$." This rearrangement creates an expression with a new appearance, but does not make any progress towards the solution. Students are often unable to make the distinction between superficial changes and deeper, structural differences. For example, novices and experts have been found to sort math problems differently (Schoenfeld & Herrmann, 1982). While novices sorted by surface features, experts sorted by deep structural features. People learning how to write computer programs often have similar troubles, because the name of a variable has more meaning to humans than to computers. It is not easy for a novice to separate the name of a variable from its contents. A novice will often copy data into a variable with a new name in preparation for an operation, incorrectly believing that the variable's name affects what the program does.

Equivalency occurs when a single state in a problem can be expressed by multiple arrangements of objects. Not being able to see through the surface features to the underlying meaning creates a layer of indirection between the solver and the core of a problem. This is a potential source of difficulty that has not been addressed in previous problem solving research.

Equivalence as an Insight

Problems can be divided into two types, insight and non-insight, with equivalence playing a role mainly in non-insight problems. An *insight* problem is characterized by an "a-ha" moment in which the solution to a problem is suddenly discovered. Typically a long period of little or no progress precedes the insight. After the insight, the goal is rapidly achieved (Metcalfe, 1986).

A *non-insight* problem is characterized by steady progress towards the goal. Non-insight problems can be further divided into "move" and "change" problems (Hayes & Simon, 1974; Simon & Hayes, 1976). Objects are moved to new locations in move problems. In change problems, the objects are stationary, but their attributes are changed. Both problem types have several objects that begin in a certain configuration, and are manipulated step-by-step to match a goal configuration. Often there are multiple groups that classify the objects. The objects can be treated individually, or as interchangeable members of their group.

However, equivalency can be thought of as a kind of insight. The equivalency in a problem may not be understood at first, but may be discovered somewhere in the middle of a solution. This change is less drastic than completely reorganizing an internal representation. It is a change that may only update a representation, but in such a way that the representation becomes more tractable.

The purpose of this thesis is to demonstrate that move and change problems are more difficult to solve if equivalency is not fully grasped. A brief review of what is known about difficulty in solving move and change problems is next. The review will discuss why memory demands need to be considered, how rules and representations matter, and then how the layout of objects can contribute to difficulty.

Many well-formed problems at their core can be described as set of states and the moves or connections between them. This network or graph representation of a problem is known as a *problem space* (Newell & Simon, 1972). The start state, goal state, and all possible intermediate states are nodes in a graph structure. For an example, see Figure B1. The rules of a problem define how moves between states can be made. Moves are represented as lines connecting the states.

Memory Demands

The size of a problem space is a basic measure of a problem's difficulty. Problems with larger problem spaces generally require more effort to solve (Kotovsky & Simon, 1990).

One reason a larger problem space makes problems difficult is that problems that require longer solutions increase memory demands. Remembering which states have already been visited is a basic strategy for finding a solution (Atwood & Polson, 1976). Problem solvers prefer not to return to previously visited states – an anti-looping strategy (Simon & Reed, 1976). Most move and change problems have linear problem spaces and solution paths that do not require solvers to revisit states to find their solution. They can be solved by visiting each state along the solution path just once. The more states a problem solver visits, the more states have to be maintained in memory to inform the anti-looping strategy.

If equivalent states are viewed as distinct states, memory demands are increased even more. Consider playing a game of Blackjack (an idealized version where counting cards does not help). If the player is dealt a six and a ten, it does not matter in which order the cards were dealt. It only matters what their sum is. The experienced player knows one rule for hitting or staying with these cards. Confused players may try to remember two rules: one for ten followed by six, and one for six followed by ten. The memory requirements have doubled for the confused players, but the additional rule does not provide them any advantage.

Rules and Representations

The rules of a problem have a clear impact on difficulty. Increasing the number of rules generally increases a problem's difficulty. The way the rules are written can also have an effect, because the description of a problem's rules affects the mental representation that is constructed and used to solve the problem (Hayes & Simon, 1974). For example, past research has investigated difficulty stemming from the complexity of the rules (Kotovsky, Hayes, & Simon, 1985). One experiment compared how different rules for the same problem required varying lengths of time to memorize them. Characters held globes in their hands and were able to alter the sizes of the globes in the change problem. Characters in the move problem passed the globes back and forth. Participants given the change rules required twice as much time to memorize them compared to participants given the move rules. The everyday applicability of the rules also affected how quickly participants found the solution. A rule that requires larger objects to be stacked on top of smaller objects is counter-intuitive. If a rule was inconsistent with everyday experience, it was harder to apply, and slowed down the problem solving process.

Zhang and Norman (1994) emphasize the need to consider the internal and external representations as a combined system. Rules that are not reinforced by the problem presentation require additional effort to be maintained in memory (Zhang & Norman, 1994).

The way the textual description of the rules is written also affects how they are interpreted and represented internally (Hayes & Simon, 1974).

A computerized presentation of a problem allows for manipulating not just the appearance, but also how people interact with a problem. The active elements on the screen can be easily varied. These elements are an external representation of the operations defined by the rules. Changing these elements can influence the internal representations used by a problem solver.

Knowing how often and what kind of mistakes are being made can reveal more about the problem solving process than just looking at the correct moves that were made. Knowles and Delaney (2005) investigated the role mistakes play in problem solving difficulty. People are more likely to try a move that violates the rules in problems with complex rules. If participants are treating objects individually and are not taking advantage of the equivalency in a problem, they perceive that they have more unique states to choose from. Eliminating equivalent states allows for a smaller set of candidate states. By keeping the candidate set in working-memory, it is more likely that all choices may be accurately evaluated (Atwood & Polson, 1976).

Layout of Objects

The spatial arrangement of objects can convey meaning in a problem without using words. Relationships can be depicted by placing objects close together. An organized and uncluttered display is also easier to process perceptually. Move problems lend themselves more easily to a spatially distinct layout than do change problems. Move problems have also been found to be generally easier than change problems (Hayes & Simon, 1974; Simon

& Hayes, 1976).

Kotovsky et al. (1985) argued that move rules are easier to visualize than change rules, partially because spatial information is easier to remember. In one experiment from this study, participants had to evaluate the legality of problem states that were briefly flashed before their eyes. Among the correct responses, participants who were given the move problem responded in less time that those given the change problem. The ease of imageability made it easier for participants to match the rules to the situations presented to them.

Move problems also provide stronger external memory aids. Kotovsky et al. (1985) argued that the external or physical representation of move problems reduces the memory load on participants. Change problems require more information to be maintained internally. This additional memory requirement made it harder for their participants to execute the mental operations necessary to apply the rules.

Current Experiments

The effects of equivalence were investigated by manipulating how objects were selected in three experiments. In all experiments, the objects in the problems were divided into two groups. The rules of the problems used the groups to define how the participants were allowed to interact with the objects. Objects within a group were interchangeable. Two selection orders were tested in each experiment: The *open* version allowed participants to directly select the objects in an order of their choosing, while the *fixed* version only allowed the choice between the two groups, with a predetermined order of particular objects within each group. Fixing the selection order was predicted to make the problem easier to solve by blocking equivalency and helping participants focus on the rules. Consequently, participants given the open version were expected to revisit equivalent states more often. The open version leaves it to the solvers to figure out the equivalency of the objects. The problem space may seem to be larger before equivalency is understood.

Experiment 2 also looked at how changes in the appearance of objects may affect equivalency. Two types of appearances were compared: *identical* and *distinct.* The descriptions of the rules were varied in Experiment 2 along with the changes in appearance. The individuality of the objects was emphasized in the *identical* groups' descriptions to try to influence how participants built their mental models. Problems in which the objects in each group shared the same identical appearance were predicted to be easier because solvers would recognize equivalent states more easily. Problems with distinct object appearances were predicted to be more difficult because people would treat each piece individually instead of treating them interchangeably.

Experiment 3 also looked at the spatial layout of objects. The move and change layouts were compared. The move version was predicted to enhance the perception of interchangeable objects within the two groups. People using the interchangeable perception were expected to switch to a more efficient strategy, making the problem easier to solve.

To preview, effects of equivalence were found between the open and fixed selection orders. People given the fixed versions consistently solved the problems in fewer moves than people given the open versions. Reliable differences were not found when comparing appearance versions and layout versions.

CHAPTER II

EXPERIMENT 1

The goal of the first experiment was to see if the presence of equivalent options for each state in the problem would increase the number of moves required to solve the problem compared to a version of the problem with only one possible configuration per state. In the problems used in this study, there is no direct benefit from revisiting past states. If people understand the problem equally well in both conditions, then there should be no additional visits to equivalent states.

Method

Participants. A total of 102 undergraduate students from the University of North Carolina at Greensboro participated in order to partially fulfill a course requirement. Two participants withdrew because they became frustrated and were not willing to complete the problem. Data from these two participants were excluded from all analyses. Twenty five participants did not finish within the time limit. This left 77 participants: 38 in the Open group and 39 in the Fixed group.

Problem and Conditions. Participants were asked to solve the Grance and Zurkistan problem. This problem depicted a fictional war between the two countries Grance and Zurkistan. Each country had four cities. The cities in Grance were displayed in the color red, and the cities in Zurkistan were displayed in the color blue. Each country had three regular cities and one capital. All cities in both countries began intact. Participants were asked to alternately destroy and rebuild the cities according to the rules shown in Table A1.

The problem was solved when all six regular cities were destroyed. Grance and Zurkistan is a new isomorph of "Hobbits and Orcs" and other "river crossing" problems (Greeno, 1974; Reed, Ernst, & Banerji, 1974; Thomas, 1974).

The problem was presented on a typical Windows personal computer. Participants used the mouse to interact with the problem's graphical user interface. The Grance and Zurkistan problem and its corresponding software was written by Andrew LeRoux.

Two variations of Grance and Zurkistan were compared to each other. Half of the participants were randomly assigned to the *Open* variation, and the other half to the *Fixed* variation. In the Open group (shown in Figure B2), participants were allowed to select the cities in any order they chose, as long as the selected cities did not violate the rules. In the Fixed group (shown in Figure B4), two paths were shown, one in each country. The paths ran through the four cities in the country. Fixed group participants were forced to select cities in order along these paths. The paths were described as highways through each country that were used by the armies. As an army advanced, it destroyed the next city; after an army retreated, the cities on the path were rebuilt. Both variations had the same problem space and solution path, as seen in Figure B3. The Fixed version of the problem limits the possible moves that can be selected.

Procedure. Participants were trained on the rules and how to use the program before starting their problem solving attempt. The first screen the participants saw showed a map of the fictional countries with the title, "The Great War of Grance and Zurkistan." The experimenter advanced to a screen of text with the background story and the four rules. Participants were asked to memorize the rules at their own pace. They then stated the rules orally to the experimenter without looking at the screen to demonstrate they had memorized them. Participants could refer back to the screen if they make a mistake, but were then required to start over until they could repeat all four rules without error.

The next several screens gave the participants practice with how to use the program and how the rules work in the problem. Advancing through the training screens was self-paced, with the experimenter watching to make sure the participant did what was described in the text. The full set of training screens is included in Appendix D. After the training and before the actual problem solving attempt, the participant was asked to repeat the rules again and state the goal of the problem. The experimenter reminded the participant what they missed if the participant forgot a rule or the goal. This was done to minimize difficulty coming from learning the rules and let the participant focus on choosing the best moves.

After having a chance to ask any final questions, the participant clicked the "Begin" button to start solving the problem. For each move, participants clicked on cities to select them, and then clicked a button labeled either "Destroy" or "Rebuild." The cities were highlighted in yellow on the screen when they were selected. A message popped up on the screen when a rule was broken. Rules 1 through 3 were enforced after the participant pressed the "Destroy" or "Rebuild" button. Rule 4 was enforced immediately if the participant attempted to highlight more than two cities.

The experimenter used a stopwatch to enforce a 20 min time limit. Pilot studies found that most people were able to solve the Grance and Zurkistan problem in 10 to 15 min. The time limit was chosen to keep the entire session under 30 min. The participants were not told there was a time limit, but they knew they had signed up for a 30 min experiment. If a participant did not solve the problem after 20 min, the experimenter offered hints so the participant could finish the problem. Hints were offered so that participants did not leave the experiment feeling bad that they could not solve the problem. Data from those given hints were only included in the analysis of solvers versus non-solvers. Their data was excluded from all other analyses. A stopwatch was used so the participant could not see the timer and did not feel external time pressure.

Results and Discussion

Solution Rates. The first question was whether people in the Open group solved the problem at a different rate than people in the Fixed group. The overall proportion of non-solvers was 0.25. In the Open group, 9 of 47 participants did not solve the problem (19%). For the Fixed group, 16 of the 55 participants did not solve the problem (29%). A chi-squared test did not result in a significant difference in solution rates between the Open and Fixed groups, $\chi^2(1) = 1.35, p = .24$. The proportion of participants who were able to solve the problems did not reveal differences in difficulty.

Legal Moves. Legal moves are made when a participant destroys or rebuilds cities while adhering to the rules. A legal move can move forward toward the solution or backwards toward the starting point. The shortest path solution to the Grance and Zurkistan problem requires 11 legal moves (see Figure B3). Few participants took the shortest path: 2.6% in Experiment 1, 6.5% in Experiment 2, and 3.0% in Experiment 3.

Counting the number of moves made to solve a problem results in a distribution of move counts across participants that does not follow the Normal curve. Observations that count the number of times something occurs have a minimum of zero, and an unbounded maximum. The distribution of counts is usually positively skewed. The Poisson distribution provides a curve that fits count data better than the Normal curve. Comparing test statistics to a Poisson distribution is more appropriate for count data (Ramsey & Schafer, 1997). Figure B5 shows the observed distribution of legal moves generally matches the Poisson shape.

A generalized linear model with legal moves as the outcome variable and condition as the predictor was created. Within the generalized linear model, a logarithmic link function was used. Using a logarithmic link function between a count outcome variable and non-count predictors makes it easier to find linear relationships between them. Legal moves were transformed by subtracting one less than the minimum number of moves required to solve the problem. This shifted the distribution of observed moves to have a minimum of one instead of a minimum of eleven. Shifting the distribution down to near zero provides a better fit with the Poisson distribution curve. Subtracting ten instead of eleven avoided zeroes for the logarithmic link function.

The condition variable was significant, $z(75) = -11.71, p < .001$. The mean number of legal moves made by the Open group was significantly larger than the mean number of legal moves made by the Fixed group (Table A2). Even though the solution path was identical in both groups, participants in the Open group made more moves to reach the goal state.

Rule Violations. Breaking a rule expresses difficulty in two ways. First, the solver is not making a move toward the goal. Second, the solver does not yet understand completely how the rules work. There are four rules in the Grance and Zurkistan problem.

The number of times any of the four rules were broken was analyzed using a model similar to the one used for legal moves. The outcome variable was the sum of all rule violations, and condition was the predictor. Table A2 gives the means and standard deviations. The rules were broken more frequently in the Open group compared to the Fixed group, $z(75) =$

$-6.63, p < .001$. Participants solving the Fixed version of the problem made fewer mistakes with the rules. Along with the increase in legal moves, the increased mistakes made by participants in the Open group shows that their version of the problem was harder to solve.

Generation Rate. At each step in the solution process, a participant needs to create a set of candidate moves and select one from this set of candidate moves. Legal moves and rule violating moves can be added to the candidate set at each step. If there is a fixed probability of generating a move that violates the rules, participants who make more legal moves will also commit more rule violations (Knowles & Delaney, 2005). A difference in rule violations could then be explained by a difference in legal moves. However, if the relationship between legal moves and rule violations varies by group, then the generation of violations cannot be fully explained by differences in legal moves. Simple linear regression was used to analyze the relationship between legal moves and rule violations. The best fitting regression equation was:

$$\text{violations} = 9.04 + (0.30 \times \text{legal}) \tag{1}$$

The coefficient for the legal term was significant, $t(75) = 7.85, p < .001$. There was a strong, positive relationship between legal moves and rule violations.

Another linear regression was conducted with Condition added as a predictor. The Open group was coded as zero, and the Fixed group was coded as one. The best fitting regression equation was:

$$\text{violations} = 10.95 + (0.29 \times \text{legal}) - (2.76 \times \text{condition}). \tag{2}$$

The coefficient for legal moves remained significant, $t(75) = 7.27, p < .001$. The coefficient for condition was not significant, $t(75) = -1.27, p = .21$. Legal moves are a good predictor of rule violations, but the condition variable did not improve the regression model.

Next, a third linear regression was conducted that included legal moves, condition, and the interaction between legal moves and condition. The coefficient of the interaction term was significant, $t(75) = 4.59, p < .001$, indicating a need to regress rule violations on legal moves separately in each condition. The regression equation for the Open group was:

$$\text{violations} = 14.49 + (0.05 \times \text{legal}) \tag{3}$$

The R^2 value was 0.34, with $t(36) = 4.28, p < .001$ for the legal coefficient. In the Fixed group, the regression equation was:

$$\text{violations} = -2.37 + (0.62 \times \text{legal}) \tag{4}$$

The R^2 value was 0.79, and $t(37) = 11.81, p < .001$ for the coefficient. Figure B6 plots rule violations against legal moves for each condition. There was a stronger relationship between legal and rule violations in the Fixed group. The Open group had more legal moves to choose from than the Fixed group. Participants in the Fixed group were more restricted in the options given to them. Participants in the Open group tended to increase their ratio of legal moves to rule violations by making more legal moves that did not directly advance them towards the goal.

Order Errors. The order was varied between conditions by letting participants in the

Open group select the cities in an order of their choosing, where as the Fixed group was required to destroy and rebuild cities in order along the paths depicted on the screen (Figure B4). An error message was displayed on the screen when a participant attempted to destroy or rebuild a city out of order. Order errors were only possible in the Fixed group. Out of the 39 participants in the Fixed condition who completed the problem, 30 committed an order error. A total of 117 order errors were committed, 65 in the forward direction (destroying cities) and 52 in the backwards direction (rebuilding cities). A t-test showed the difference between the directions was not significant ($t < 1$).

Reversals. Undoing the previous move to go back to the state that was just visited is a reversal. A reversal is executed by a pair of legal moves, one forward and one backward. In the current problem, there were two kinds of reversals. Going back to the same configuration of destroyed cities is an *exact* reversal. Returning to a different selection of cities, but maintaining the same number of destroyed cities in each country, is an *equivalent* reversal (Figure B7). An example of an equivalent reversal would be destroying the first and second cities in Zurkistan, then rebuilding both cities, and finally destroying the first and third cities. To advance along the solution path, it does not matter which specific cities are destroyed, as long as the count of destroyed cities in each country satisfy the rules. An exact reversal would be destroying, rebuilding, and destroying the same city. In the Open group, exact and equivalent reversals are possible. In the Fixed group, only exact reversals are possible since the participant does not have a choice of which specific city to select.

People in the Open group were predicted to make more total reversals than people in the Fixed group. Poisson-based regression was used to analyze the reversal counts. The total reversals include equivalent and exact reversals. All reversals were first regressed on condition. The mean for all reversals was significantly higher in the Open condition compared to the Fixed condition, $z(75) = -11.67, p < .001$. There was no difference found between the Open and Fixed groups, $z < 1$, when examining exact reversals only. Participants in both conditions returned to the exact same state at similar rates. However, people in the Open group made additional reversals back to equivalent states. This indicates that participants did not completely understand the equivalency of some states. The available additional configurations of previously visited states were explored as if they were new states.

State Visit Counts. The state visit counts are shown here to allow comparison with previous research. The same states that were difficult in the Grance and Zurkistan problem were the ones that were difficult in earlier river-crossing studies (Greeno, 1974; Thomas, 1974). Figure B8 shows the mean number of visits to each state observed in Experiment 1. The x-axis shows the solution path progressing from left to right. These states are the same as those shown in Figure B3.

Summary and Conclusions. Fixing the order helped people to solve the problem as indicated by participants making fewer legal moves and rule violations. People given the fixed version only had two locations to choose from - the current heads of the paths. Restricting the possible selections blocked the possibility of trying equivalent combinations of cities. Those given the open version could choose any of the cities. They were treating each city as a unique object. There were a total of six objects for them to consider, excepting the two capital cities that were excluded by the first rule. This increased the perceived size of the problem space for those given the open version.

The cover story also implicitly emphasized the distinctiveness of the objects by describing a war. In a war, not all cities are equal. Some may have strategic advantages over others. However, only the total count of destroyed cities in each country is important to the rules of this problem. Real world knowledge may have hindered the discovery of the solution. If the equivalency in this problem was completely understood, then the solvers given the open version would have treated the cities as two groups of three interchangeable objects.

Participants in the Open condition may not have realized they were going back to equivalent states. The same number of exact reversals were made in each condition. Additional equivalent reversals were made in the Open condition compared to the Fixed condition. People given the open version may have believed the problem space was larger than it actually was. Exploring these extra states would be a useful strategy if they were not logically equivalent to states that had already been visited. The equivalence may not have been understood because people treated the cities as six unique objects, instead of two groups of three equivalent objects.

CHAPTER III

EXPERIMENT 2

In Grance and Zurkistan, each city is drawn on the screen with a different shape, and in its own location on the map. These perceptual differences may cause the solver to believe that the specific cities that are chosen is important in solving the problem, when in fact this is irrelevant. The first goal of Experiment 2 was to test whether perceptually distinct appearances of objects would make it more difficult to understand equivalence. To test the effect of appearances, we turned to an isomorph of Grance and Zurkistan, the Hobbits and Orcs problem.

In Hobbits and Orcs, the hobbits are typically all exactly the same, and the orcs are all exactly the same. It may be easier for the solver to understand that any hobbit or orc is just the same as any other hobbit or orc. If the appearance of the hobbits and orcs is made distinct, it was hypothesized that solvers might not grasp their equivalency as quickly as people who are shown identical hobbits and orcs. Figure B9 shows the character icons used in Experiment 2.

A second goal of Experiment 2 was to replicate the selection order effect from Experiment 1. Participants may consider a different selection of the same kind of traveler a new state. By removing this choice, some of the complexity of the problem may be reduced, resulting in fewer moves to solve in the Fixed condition compared to the Open condition.

Method

Participants. A total of 153 undergraduate students from the University of North Carolina at Greensboro participated in order to partially fulfill a course requirement. Data from 18 people were excluded from the analysis. One participant did not speak English well and was given extra help to understand and complete the problem. Another participant solved the problem during the practice before the actual experimental problem began. One participant did not want to finish the problem and withdrew. An additional 8 participants did not finish the problem within the 20 min time limit. Finally, 3 of the 18 excluded participants had solved the problem before. These exclusions left data from 139 students that were included in the following analyses.

Design. The experiment manipulated the order in which the travelers were chosen and the appearance of the travelers. This made a 2 Order (Open versus Fixed) by 2 Appearance (Identical versus Distinct) between-subjects design. Participants were randomly assigned to one of the four conditions upon entering the lab.

Problem and Conditions. The Hobbits and Orcs problem has six travelers: three hobbits and three orcs. These travelers come to a river that they must cross. There is one boat for them to use. All travelers begin on the left side of the river, and the goal is to safely transport all six to the right side of the river. The three rules that govern how the travelers may move are shown in Table A1.

The problem was presented to the participants on a typical Windows personal computer. Participants used the mouse to interact with the program. I wrote the software implementation of Hobbits and Orcs used in this experiment.

The two appearance variations altered the images used to represent the travelers on the computer screen (see Figure B9). In the *Identical* condition, all of the hobbits used the same image and all of the orcs used the same image. In the *Distinct* condition, each hobbit and each orc was represented by a distinct image. Each traveler was also labeled with a unique name to further emphasize his distinctiveness. The names were always displayed above each traveler's image.

The order of selecting travelers was enforced by restricting how travelers could be added to and removed from the boat. In the *Open* group, any hobbit or orc that is on the same side of the river as the boat could be directly clicked on and added to the boat. The *Fixed* group was given three buttons on each side of the river. The buttons were labeled "Add Hobbit," "Add Orc," and "Remove." Figure B10 shows a screenshot of the Fixed group's program. Travelers were added to the boat by clicking on these buttons instead of by clicking directly on the travelers, which removed the choice of which specific traveler was added to the boat.

Procedure. Participants first memorized the rules using the same procedure as Experiment 1. The next screen shown to the participants was a practice screen. It had the same appearance as the actual problem display, with the exception of the word "practice" written down the middle of the screen in large letters. The experimenter read aloud a script that described the problem, and instructed the participant to perform four practice moves. The practice moves were designed to familiarize the participant with how to move the travelers around. The experimenter waited until the participant had completed the current instruction before moving on to the next. The scripts in the Open conditions told the participant to click directly on the travelers. In the Fixed conditions, the scripts described using the buttons to add and remove travelers from the boat. The scripts also differed in how they referred to the travelers. In the Identical conditions, travelers were identified using indefinite articles (“a hobbit” or “an orc”). The Distinct conditions used the travelers’ names. The scripts are included in Appendix C.

After completing the training, participants began their problem solving attempt. As in Experiment 1, a 20 min time limit was enforced.

After the problem, the experimenter asked the participant if they had seen or solved the problem, or other problems like it before. It is possible that some participants may have been exposed to the Hobbits and Orcs problem (or some variation) in the past. The data from these participants was excluded from the following analyses.

Results and Discussion

Solution Rates. As in Experiment 1, the analysis of difficulty began by seeing if people solved the problem more often in one condition compared to the other. A total of 8 people ran out of time attempting to solve the problem (5.4%), while 139 participants solved it. Table A4 contains the proportions broken down by condition. A chi-squared test did not find a significant difference in solution rates, $\chi^2(3) = 1.20, p = .75$. This simple analysis did not provide evidence for differences in difficulty.

Legal Moves. A legal move in Experiment 2 was defined in the same way as Experiment 1: a move that did not break any of the rules. Means and standard deviations for each condition are given in Table A5. Move counts were analyzed in the same way as in Experiment 1, by using a Poisson distribution and a generalized linear model with a logarithmic link function. The Poisson shape of the observed legal moves distribution is shown in Figure B11. Legal moves were shifted down again by subtracting ten moves - one less than the minimum number of moves required to solve the problem. The marginal mean for legal moves in the Open versions was 17.68 ($SD = 16.11$), and the marginal mean for the Fixed versions was 14.68 ($SD = 11.71$). There was a main effect of selection order, $z(137) = -4.38, p < .001$. The main effect of appearance was not significant, $z < 1$. The marginal mean for the Identical versions was 16.18 ($SD = 13.26$), and the marginal mean for the Distinct versions was 16.35 ($SD = 15.42$). The interaction was also not significant, $z(137) = 1.19, p = .23$.

As in Experiment 1, the main effect of selection order was highly significant. People given the Fixed version solved the problem with fewer legal moves. However, controlling the appearance of the hobbits and orcs did not affect legal moves.

Rule Violations. Were the rules harder to follow in some conditions compared to the others? The number of times any of the rules were broken were summed for each participant, and are given in Table A5. A generalized linear model using the Poisson distribution and a logarithmic link function was used again to see if the rule violations were affected by selection order or appearance. Three models were created. The total number of rule violations was regressed on selection order only, then on appearance only, and then on both and interaction of selection order and appearance. In the first two models, the coefficients for selection order and appearance were not significant, $z < 1$. The third regression also did not yield any significant coefficients. The main effect terms had $z < 1$. The interaction term had $z(135) = 1.43, p = .15$. Unlike Experiment 1, the number of rule violations was fairly consistent across all conditions.

Generation Rate. The relationship between legal moves and rule violations was examined in the same way as in Experiment 1. A difference in the relationship between legal moves and rule violations could show that the conditions are affecting how people think through the problems. The first regression model predicted rule violations with legal moves. The best fitting equation was:

$$\text{violations} = -3.08 + (0.69 \times \text{legal}) \tag{5}$$

The coefficient 0.69 was significant, $t(137) = 15.20, p < .001$.

The next step added selection order as a predictor. The selection variable was coded as open equal to zero, fixed equal to one. The best fitting regression equation was:

$$\text{violations} = -4.33 + (0.70 \times \text{legal}) + (2.19 \times \text{selection}) \tag{6}$$

The coefficient for legal moves was significant, $t(137) = 15.40, p < .001$. The coefficient for the selection variable approached significance, $t(137) = 1.70, p = .09$.

The interaction between legal moves and selection order was added next. The coefficient for the interaction term was significant, $t(137) = 4.22, p < .001$. Follow up regressions within the Open and Fixed groups came next. The best fitting regression equation for the Open group was:

$$\text{violations} = -0.88 + (0.57 \times \text{legal}) \tag{7}$$

The legal coefficient was significant, $t(71) = 12.14, p < .001$. The regression equation for the Fixed group was:

$$\text{violations} = -8.61 + (0.96 \times \text{legal}) \tag{8}$$

The coefficient for the legal term was significant again, $t(64) = 11.67, p < .001$. The larger coefficient shows that rule violations were made at a higher rate in the Fixed group (Figure B12). Selection order affected the generation of rule violations.

Did the appearance manipulation affect the generation rate? Identical appearances was coded as zero, and distinct appearances was coded as one. This term was added as a predictor with legal moves, resulting in:

$$\text{violations} = -3.11 + (0.69 \times \text{legal}) + (0.074 \times \text{appearance}) \tag{9}$$

The coefficient for the legal term remained significant, $t(137) = 15.14, p < .001$. However, the appearance coefficient was not significant, $t < 1$. The interaction of legal moves and distinct appearance was added next. The interaction term was not significant, $t(137) = 1.29, p = .20$. Changing the appearance of the travelers did not affect the rate of rule violations.

Reversals. Similar to the Grance and Zurkistan problem in Experiment 1, the Hobbits and Orcs problems affords two kinds of reversals, exact and equivalent. The selection order was expected to operate on reversals the same way in Hobbits and Orcs as it had in the Grance and Zurkistan problem. Distinct appearances for the characters were predicted to make it harder to recognize equivalent states and increase the number of reversals. When an open selection order was combined with distinct appearances, it was predicted that they would have an additive effect in the same direction. Means and standard deviations for the reversals are shown in Table A6. Poisson-based regression models were used to analyze equivalent and exact reversals.

First, all reversals were regressed on selection order. The marginal mean for reversals in the Open group was 10.93 ($SD = 7.81$), and the marginal mean for the Fixed group was 8.68 ($SD = 6.40$). The coefficient of the selection order term was significant, $z(137) = -4.21, p < .001$. Fixing the selection order decreased the number of reversals.

Next, reversals were regressed on appearance. The marginal mean for the Identical group was 9.62 ($SD = 6.37$), and the marginal mean for the Distinct group was 10.16 ($SD = 8.21$). The coefficient for the appearance term was not significant, $z(137) = 1.01, p = .31$. Appearance did not reliably change how many reversals were made.

There was also no significant interaction, $z < 1$. When considering all reversals, only selection order had a significant effect. People in the Fixed group made fewer reversals than those in the Open group.

Conclusions. One of two main hypotheses for Experiment 2 was confirmed. As in Experiment 1, there were fewer legal moves made in the Fixed groups compared to the Open groups. There were also fewer reversals in the Fixed groups compared to the Open groups. The option to choose equivalent states was blocked in the Fixed versions. This reduced the perceived size of the problem space compared to the Open version. Participants in the Open groups explored the additional states without realizing that they were equivalent to other states that had already been visited.

However, the number of rule violations did not differ between conditions, as it had in Experiment 1. There are three rules to memorize in the Hobbits and Orcs problem compared to four rules to memorize for Grance and Zurkistan. The Hobbits and Orcs rules are also easier to understand because they correspond more closely to real-world experience (Kotovsky et al., 1985). Compliance with the rules may have been near ceiling in Experiment 2. The more difficult rules in Experiment 1 may have allowed for differences due to selection order to be observed in the data.

The second hypothesis expected that the characters with distinct appearances would make it more difficult for the participants to understand the equivalency of the two groups of characters. However, no significant differences were found between the Identical and Distinct versions of the problem. The change in appearances also did not affect the number of reversals. It is possible that the hobbits and orcs in the Distinct version were not different enough from each other to cause a problem solver to treat them as individuals. However, the different shapes of the cities on the map in Experiment 1 were less different from each other than the icons used for the characters in Experiment 2. If the appearance of objects was affecting equivalency in Experiment 1, the changes in appearances in Experiment 2 should have caused differences in the number of moves made. This lack of evidence leads away from an explanation based on appearance.

If how the objects look is not important, perhaps their spatial arrangement on the screen impacts how equivalency is understood. The hobbits and orcs were moved back and forth across the river in Experiment 2. This movement could have been a stronger cue to equivalency that overwhelmed changes in appearance.

CHAPTER IV

EXPERIMENT 3

In the Hobbits and Orcs problem, the travelers' icons move back and forth across the river. In Grance and Zurkistan, the cities stay in place, and their appearance is altered when they are destroyed. It is possible that the more spatially distinct Hobbits and Orcs are easier to keep track of than the stationary cities in Grance and Zurkistan. Experiment 3 investigated this spatial hypothesis with Grance and Zurkistan, and again attempted to replicate the selection order finding in Experiment 1.

The version of Grance and Zurkistan used in Experiment 1 can be classified as a "change" problem. The cities do not change position. Different problem states were defined by destroying or rebuilding the cities. A "move" problem defines new states by moving elements to new positions. A move isomorph of Grance and Zurkistan was devised that moved cities to either side of the screen depending on whether they were intact or destroyed. Previous research has found change problems to be harder than move problems (Simon & Hayes, 1976; Kotovsky et al., 1985).

It was hypothesized that the new "move" isomorph of Grance and Zurkistan would require fewer moves to solve than the original "change" version. The open versus fixed selection order manipulation was included to replicate the results from the first two experiments, and to see if it interacted with the move versus change versions. The fixed version was expected to be easier than the open version for both the change and move versions.

Method

Participants. The participants were 82 undergraduate students from the University of North Carolina at Greensboro. Participants received course credit for participating. One participant was excluded because the participant had solved the problem in the past. Fifteen other participants did not solve the problem within the specified time limit. After these exclusions, data from 66 participants were used in the following analyses.

Design. Experiment 3 aimed to see how the difficulty of the Grance and Zurkistan problem was affected when it was reconfigured from a "change" problem to a "move" problem. This is the *Isomorph* variable. The *Open* versus *Fixed* selection order manipulation was identical to Experiment 1. This resulted in a 2 Isomorph (Change versus Move) x 2 Selection Order (Open versus Fixed) between-subjects design.

Problem. The Grance and Zurkistan problem was used again in Experiment 3. The new "move" version of Grance and Zurkistan moved cities to different locations on the screen. The screen for the Move version was split horizontally into two halves. The cities began intact, and were displayed on the left side of the screen. Cities were moved to the right side of the screen to indicate that they were destroyed, and back to the left side when they were rebuilt. At the top of both sides were labels describing the state of the cities, "Intact Cities" on the left and "Destroyed Cities" on the right. A city kept its same position relative to the other cities when it was moved. Figure B13 shows two destroyed cities in Zurkistan moved to the right side of the screen.

Procedure. The procedure was the same as Experiment 1, except for a few modifications to the training text read by participants. The updated text described how cities would move to the left and right sides of the screen depending on their state.

Results and Discussion

Solution Rate. Overall, 15 out of 85 people ran out of time while attempting to solve the problem (18%). The individual solution counts for each cell are given in Table A7. A chi-squared test did not result in a significant difference in proportions, $\chi^2(3) = 4.42, p = .22$. As in the first two experiments, the solution rates did not reveal any obvious differences in difficulty.

Legal Moves. Figure B14 shows the distribution of legal moves. A bulge at the right side of the histogram can be seen in Figure B14. This may have been due to people who were not following directions, and were not engaged in trying to figure out the problem. Data from participants who made more than 140 total moves were excluded from this analysis. This dropped 1 participant from the Move-Fixed condition, and 3 participants from the Change-Fixed condition.

Analyzing legal moves reveals if one or more of the conditions required fewer moves to find a solution. Means and standard deviations for the legal moves are shown in Table A8. After subtracting ten from the legal move counts to shift the minimum to near zero, Poisson based, generalized linear models with logarithmic link functions were used again. The factors were Selection Order (Open versus Fixed) and Isomorph (Move versus Change).

First the effect of selection order on legal moves was tested. The mean number of legal moves for the Open groups was 23.24 ($SD = 16.89$), and the mean for the Fixed groups was 21.59 ($SD = 13.53$). When legal moves were regressed on selection order, the coefficient was not significant, $z(64) = -1.41, p = .16$. Changing the selection order did not result in a meaningful difference in legal moves.

Did changing the Isomorph type affect legal moves? The marginal mean for the Change conditions was 26.00 (SD = 22.60), and the marginal mean for the Move conditions was 23.43 (SD = 14.78). When legal moves were regressed on Isomorph, the coefficient for the term was not significant, $z(64) = -1.15, p = .25$. There was not a reliable difference in the number of legal moves made by participants across the Change and Move isomorphs.

An interaction model was created next. The interaction of Isomorph type and Selection Order was significant, $z(62) = 2.01, p < .05$. Follow-up regressions within each Isomorph type were done next. Within the Move isomorph, selection order did not have a significant effect, $z < 1$. The mean for legal moves in the Move-Open group was 21.67 (SD = 12.78), and the mean for the Move-Fixed group was 22.21 (SD = 14.46). There was a significant difference between the Selection Order means within the Change Isomorph type, $z(29) = -2.49, p < .05$. The mean for Change-Open was 25.31 (SD = 21.42) and the mean for Change-Fixed was 21.00 (SD = 13.09). The Change version of the problem is the same problem that was used in Experiment 1. A similar reduction in legal moves was found here in the Fixed group.

Rule Violations. Did people break the rules more often in the Open selection order group compared to the Fixed selection order group? The number of rule violations for each of the four rules were summed together for each participant. The same outlier criteria and Poisson-based regression approach as the legal move analysis was used here. The marginal mean for the Open groups was 21.59 (SD = 13.37), and the marginal mean for the Fixed groups was 22.79 (SD = 11.10). When rule violations were regressed on selection order, the coefficient was not significant, $z(64) = -1.03, p = .30$. Fixing the selection order did not have a reliable effect on how often the rules were broken.

Isomorph type did not affect rule violations either. The marginal mean for the Change groups was 21.42 ($SD = 12.55$), and the marginal mean for the Move groups was 22.74 ($SD = 12.32$). The regression coefficient was not significant, $z(64) = 1.14, p = .25$. The interaction of Selection Order and Isomorph type also did not yield a significant result, $z < 1$. Altering the locations of cities on the screen did not affect how often the rules were broken.

Generation Rate. The rate at which rule violations were made was examined again in Experiment 3. Was the relationship between legal moves and rule violations stronger in any of the conditions? Linear regression was used to analyze this relationship. The first model regressed rule violations on legal moves for all participants. The best fitting equation was:

$$\text{violations} = -2.01 + (0.82 \times \text{legal}) \tag{10}$$

The coefficient was significant, $t(68) = 13.98, p < .001$. Overall, a strong, positive relationship was found between the number of legal moves and the number of rule violations.

The effect of selection order was examined next. Selection order was again coded as open equal to zero, and fixed equal to one. The best fitting regression equation was:

$$\text{violations} = -4.55 + (0.80 \times \text{legal}) + (6.92 \times \text{fixed}) \tag{11}$$

The Fixed coefficient was significant, $t(68) = 2.02, p < .05$. The legal coefficient remained significant, $t(68) = 14.59, p < .001$. The interaction of legal moves and fixed selection was added next. The interaction term approached significance, $t(68) = 1.90, p = .06$. Regressions for each selection order were conducted next. In the Open group, the best fitting regression equation was:

$$\text{violations} = -0.73 + (0.69 \times \text{legal}) \tag{12}$$

The coefficient was significant, $t(35) = 9.46, p < .001$. For the Fixed group, the resulting equation was:

$$\text{violations} = -1.03 + (0.89 \times \text{legal}) \tag{13}$$

The coefficient was significant, $t(31) = 11.19, p < .001$. The rate of rule violation generation was higher in the Fixed group (Figure B12), as in Experiments 1 and 2.

Did the other factor, Isomorph type, affect the relationship between legal and rule violations? The Isomorph variable, move, was coded as change equal to zero, move equal to one. This variable was added to the initial regression model. The resulting equation was:

$$\text{violations} = -2.38 + (0.82 \times \text{legal}) + (0.60 \times \text{type}) \tag{14}$$

The coefficient for the legal term remained significant, $t(68) = 13.82, p < .001$. However, the coefficient for the type term was not significant, $t < 1$. The last model added the interaction between legal moves and isomorph type. The coefficient for the interaction was not significant, $t(68) = 1.32, p = .19$. There was no difference in generation rate found when comparing the Change and Move versions of the problem.

Order Violations. Order violations occur when a city is chosen out of order. For example, if no cities are destroyed, and the last city along the path is chosen first, that is an order violation. Order violations are only possible in the Fixed versions of the problem. There were 33 participants that were given the Fixed version. Of those 33, 28 committed at least one order violation. In the Move-Fixed condition, the mean number of order violations was 2.80 (SD = 2.43). The mean was 3.39 (SD = 3.40) in the Change-Fixed condition. A t-test showed this difference was not significant, $t < 1$. Order violations occurred at similar rates in the Change and Move versions of the problem.

Reversals. Reversals in Experiment 3 have the same meaning as they did in Experiment 1 and Experiment 2.

The selection order was expected to have an effect in Experiment 3 that was similar to its effects in Experiments 1 and 2. People in the Open groups were able to return to a state with the same number of cities destroyed in each country, while choosing a different combination of particular cities. Fixed group participants could only go back to the exact same state, and would choose to do this less often. It was expected that the reversals in the Open groups would be higher than the reversals in the Fixed groups. The marginal mean for the Open conditions was 12.49 (SD = 11.03), and the marginal mean for the Fixed condition was 10.00 (SD = 6.72). A Poisson regression revealed a significant difference in the means, $z(64) = -2.96, p < .01$. Fixing the selection order again reduced the number of reversals.

Did the Isomorph type affect reversals? It was predicted that the Move version of the problem would make the concept of equivalency easier to understand, and result in fewer reversals. The marginal mean for the Change versions was 11.97 (SD = 10.52), and the marginal mean for the Move versions was 10.89 (SD = 8.42). The coefficient for the

Isomorph was not significant, $z(64) = -1.30, p = .19$. Adding an interaction term also did not improve the model, $z < 1$. Unexpectedly, the Move version did not have an effect on reversals.

Conclusions. The only reliable difference in legal moves in Experiment 3 was found between Open and Fixed selection orders within the Change isomorph. This is the exact same effect seen in Experiment 1. The Change versus Move isomorph types did not reveal any differences in legal moves. The selection order effect was also not found in the Move isomorph.

No significant differences in rule violations were found in Experiment 3. There may not have been enough participants within the Change isomorph to replicate the effect found in Experiment 1, though the observed difference was also in the wrong direction in Experiment 3.

The effect on reversals from Experiment 1 was replicated in Experiment 3. Selection order did affect reversals. There were fewer reversals in the Fixed groups compared to the Open groups. However, changing isomorph type did not have the expected effect on reversals.

Altering Grance and Zurkistan from a "change" problem into a "move" problem may have introduced more confusion. The modifications required to make it a move problem did not clarify the goal with respect to the cover story. There was not an intuitive reason for separating the displays of intact and destroyed cities. The structure of the problem fit the requirements for move problems, but unlike Hobbits and Orcs, the cover story did not. In Hobbits and Orcs, the move nature of the problem makes intuitive sense with the cover story of crossing a river. The arbitrary rules, smaller sample size, and increased variability in Experiment 3 made it difficult to observe significant differences.

CHAPTER V

GENERAL DISCUSSION

The effect of equivalency on problem solving difficulty was investigated in three experiments. Equivalency occurs when multiple combinations of objects represent a single state. Because equivalent states are logically the same, visiting all equivalent states is not required when searching for the goal state in a problem space. An efficient solution would minimize the visits to logically equivalent states.

Two river crossing problems were used in this study: Hobbits and Orcs, and Grance and Zurkistan. These problems are isomorphic, meaning that they have the same underlying structure and the same pattern of moves will solve both problems. The problems were chosen because it was relatively easy to constrain equivalence by fixing the order in which travelers could be selected.

Selection Order

The most consistent finding from the three experiments was the benefit of fixing the order in which objects were selected. Experiment 1 used the Grance and Zurkistan problem to test how different methods of selecting cities affected how many moves were needed to solve the problem. The order of city selection was either left open or fixed in a predefined order. The Fixed version of the problem was easier, as shown by reduced numbers of legal moves, rule violations, and reversals compared to the Open version. Experiment 2 replicated the selection order finding from Experiment 1.

Experiment 3 again found the fixed version to be easier than the open version, but only within the change version of the problem. The change version in Experiment 3 was an exact replication of the problem used in Experiment 1. There was no difference found in the legal moves and rule violations made between the Open and Fixed groups within the move version.

The mechanism for fixing the order of selection in Experiment 3 was not as easy to see as it was in the first two experiments. The Move version of Grance and Zurkistan in the third experiment changed the position of cities when they were destroyed or rebuilt (Figure B13). When cities were displayed on both sides of the screen, the path through the country was not continuous and did not visually reinforce the order restriction as strongly as the Change version of the problem. Participants could easily forget this restriction and commit more order violations.

The perceived size of the problem space was larger for participants in the Open group. Fixing the order of selection allowed only one combination of destroyed and intact cities for each state. Participants in the Open group explored additional combinations by exploring additional configurations of the same equivalent state. These configurations were seen as completely new states because the solvers did not treat the cities within a country as interchangeable objects.

Appearance

People in the first experiment may have had trouble understanding the equivalence of cities because they were drawn as different shapes on a map. It was predicted in the second experiment that the distinct appearances of the hobbits and orcs would mimic the distinct shapes in Grance and Zurkistan. The distinct appearances might increase the perceived number of states to explore. By reducing the equivalence of the states, participants were expected to take more moves to solve when compared to the Identical condition.

Making each hobbit and orc have a distinct appearance did not increase the number of moves made by participants. The appearances may not matter as much as the display of separate icons on the screen. The display of separate icons may be a strong cue for people to treat the travelers as individuals.

Instead of emphasizing the distinctiveness, it may be necessary to make the characters appear more similar to each other. A future study could change the display of the Grance and Zurkistan problem so the cities were not displayed as a map. The cities could be displayed using two sets of identical icons, arranged in two lines. This arrangement should mirror the arrangement of objects in the hobbits and orcs display.

Reversals

Increased difficulty due to equivalent states was also revealed through the greater number of reversals made in the open versions compared to the fixed versions. A significant difference in reversals made was seen between Open and Fixed groups in all three experiments.

The observed differences in reversals are closely related to the differences found for legal moves. Since reversals are legal moves, it follows that people who made more legal moves must have been revisiting past states. However, reversals provide more focused evidence of difficulties understanding equivalency. A reversal is an immediate return to a previously visited state. Research on the anti-looping heuristic (Atwood & Polson, 1976; Simon & Reed, 1976) has shown that people try to avoid revisiting states. Memories of the past one or two most recently visited states can be assumed to be strong. Executing a reversal to an equivalent state shows that people prefer not to go back to exactly the same state, but do not treat equivalent states as logically identical.

In the Open group, people repeated their moves, but did not perceive them as a repetition. Even though the items chosen were equivalent, being able to choose a new combination of items made a repeat move feel different enough to try. When a problem is still new, a solver has not learned that some moves are the same as others.

In the Fixed condition, the smaller number of reversals showed that when not given a choice, people did avoid repeating moves. If all of the travelers or cities were seen as completely interchangeable, the number of reversals in the Open condition would have matched the number of reversals in the Fixed condition.

Legal Moves and Rule Violations

The relationship between legal moves and rule violations was found to be different when comparing open and fixed selection orders. The relationship was stronger in the fixed conditions. This pattern was found in all three experiments.

People have an aversion to making illegal moves. In the Open condition, there are seemingly more moves available to try than in the Fixed condition. Moves to equivalent states are legal moves; they are not eliminated simply by checking them against the rules. They must evaluated by the solver against moves made in the past and the current position along the solution path. Even though these additional moves are equivalent to other moves that may already have been tried, they feel different enough to the solver to be an appealing option.

Compared to the Open condition, the Fixed condition had fewer legal moves available to try. The rate of generating a move that violated the rules was higher in the Fixed conditions. The increased exposure to rule failures may help solvers understand the problem more quickly than those in the Open condition. The additional options in the Open condition delayed the necessary rule violations that teach solvers the most about how the problem works. Increased understanding of equivalency aids in the search for moves that are actually new. Being able to "fail fast" was beneficial to the solvers given the fixed versions.

Atwood and Polson Model

Existing models of problem solving cannot directly address differences in moves related to equivalency manipulations, but equivalency would be a fitting addition. Atwood and Polson (1976) developed a computational model that described how people solve water jugs problems (Luchins, 1942), which was later adapted to river crossing problems (Jeffries, Polson, Razran, & Atwood, 1977). The model used a three stage move selection process. Stage 1: Generate and evaluate moves. While avoiding old states, a move is selected that has the best result from a heuristic evaluation function. Stage 2: If no move was selected in Stage 1, randomly choose a move to a new state. Stage 3: If no move was selected in Stage 2, reuse the evaluations from 1 to choose the best move to an old state, or make a random move.

Incorporating equivalency could refine Atwood and Polson's model if long-term memory fallibility is also included. Atwood and Polson used perfect long-term memory as a simplifying assumption in their model. Returning to previously visited states would only occur in their model if a solver had exhausted all options and fallen back on a random selection strategy. Human memory is, of course, imperfect. If equivalency is not fully understood in a problem, additional states are added to the memory record. There are 15 distinct states in the problem space for Hobbits and Orcs and Grance and Zurkistan. If equivalence is not understood, the number of distinct states increases to 55. This makes it much harder for a solver to remember all of the states they have visited in the past. They may choose a state several times without realizing they have been there before.

Equivalency could also affect the generation of candidate moves in the model. When equivalency is understood by the solver, or is blocked by fixing the selection order of objects, there are fewer states to choose from. The increase in possible states due to unrecognized equivalency could exceed the limits of short term memory. In Atwood and Polson's model, the evaluation of states ends when short term memory is exceeded. A solver will then switch to a simpler strategy that randomly chooses new states instead of evaluating the candidate states and choosing the best option. Equivalency can be seen as a tool a solver will use to limit memory demands of a problem.

Change versus Move

Reformulating Grance and Zurkistan from a change problem into a move problem had little effect on problem solving behavior. Since changing the appearances had little effect in Experiment 2, Experiment 3 changed where the cities were located on screen in the Grance and Zurkistan problem. A variation on the Grance and Zurkistan problem kept intact cities on the left side of the screen, and destroyed cities on the right side of the screen. The equivalence of states was expected to be easier to grasp in the spatially separated "move" version.

The move version did not end up being easier than the change version. Differences in the number of moves made were not found. This could be because the Grance and Zurkistan problem was too hard. The Grance and Zurkistan problem was more difficult than the Hobbits and Orcs problem. In both Experiment 1 and Experiment 3, people required more moves to solve Grance and Zurkistan compared to the number of moves required to solve Hobbits and Orcs in Experiment 2. This is likely due to the additional complexity found in the Grance and Zurkistan rules. There are four rules in Grance and Zurkistan compared to three rules in the Hobbits and Orcs. The rules in Hobbits and Orcs also have a natural and intuitive logic to them, making them easier to remember and apply. The Grance and Zurkistan rules are somewhat arbitrary and have little in common with everyday experience. The second rule requires two cities in Zurkistan to be destroyed whenever two cities in Grance are destroyed. No reason is given for this restriction. The third rule is similarly abstract. Kotovsky et al. (1985) found evidence that increasing rule complexity made problems with the same solution more difficult. Participants may have already been solving the Grance and Zurkistan problem at ceiling levels. This low ceiling may not have left enough room for group means to diverge.

Future studies could try to vary equivalency within easier problems, or by requiring the solvers to reach some level of facility with the problem before introducing variations. Using various starting and ending configurations within the same problem has been a useful way of giving repeated exposure to the same problem without letting people memorize the solution (Anderson & Douglass, 2001).

Concepts Shared With Analogies

Finding the appropriate place to apply past problem solving knowledge to new problems is related to analogical processing. The Hobbits and Orcs problem has a particularly difficult sequence in the middle of the problem. This was seen in an earlier study that used Hobbits and Orcs (Thomas, 1974). Figure B8 has a peak at States 6 and 7 that shows the same pattern of difficulty was found in the current study. Thomas tried to alleviate the difficulty at this point by giving his participants practice on the move required to get past State 7. The practice was given before the participant attempted to solve the full problem. Similarly, the computer programs that administered the problems in the current study gave the participants thorough training on the rules of the problem prior to the problem solving attempt. This training emphasized the application of the particular rules that could lead the solver past States 6 and 7.

Thomas found that the practice did not help. People given practice on the critical moves still struggled at States 6 and 7. Most participants attempting the current problems became stuck at States 6 and 7 (Figure B8). One explanation is the participants in both studies did not have enough experience yet with the problem for the practice or training to help. They may be able to reproduce the training scenario, but not apply it when it is required in the full problem. This could also be interpreted as an analogical failure. People may have not recognized it was time to apply their practice when they arrived at State 6.

A switch from treating objects individually to treating them as interchangeable objects is similar to the "analogical shift" (Gentner, 1983). People begin a comparison by mapping the literal features to each other. When they go beyond the literal features and start comparing the relations within and between their observations, this marks a shift to more analogical reasoning. In Gentner's structure mapping framework, the relations between objects are more important to making analogies than surface features. Relationships between objects may be represented externally by a problem's rules and display, but the higher order relations between relations described by Gentner are not found in the problem. They are only discovered or understood in the internal mental representations generated by a problem solver.

The analogies made while solving Hobbits and Orcs or Grance and Zurkistan are not all from past experience with other problems. Fully understanding equivalence requires a solver to use analogies. The solver must recognize that selecting an object from one group is exactly the same as selecting any other object from the same group.

As seen in Experiment 2, changing the surface features did not affect the number of moves made by participants. People were already operating at a deeper level below the surface features. The first kind of analogical shift from surface features to deeper comparisons seemed to happen automatically.

Changing the interface features that controlled how people interacted with the problem had a reliable effect. Fixing the selection order also changed how people used the mouse to make moves. Instead of clicking directly on the object as in the open versions, people in the fixed versions clicked on buttons which then indirectly selected the specific objects. Many participants seemed to initially favor a trial-and-error strategy, rather than focusing on interpreting and applying the rules. Changing how people interact with the problem may encourage them to think about what is happening at a conceptual level. This could result in fewer repetitions and faster solutions.

Conclusions

Evidence was found for people erroneously believing some equivalent states were actually new states worth visiting during their search for a solution. This belief increased the time and effort required for them to solve the problem. The effects of distinct appearances and spatial arrangement are still believed to be meaningful influences on recognizing equivalence, but greater control over other potential sources of difficulty needs to be exerted to reduce unwanted variance in the data.

A key advantage in Experiment 2's Fixed conditions was the use of buttons to add and remove travelers from the boat instead of clicking directly on the travelers. The buttons aligned with the movement restrictions. The buttons encouraged people to focus on two groups of three objects instead of considering six objects individually. Representing the operators optimally helps people solve problems efficiently. When participants encountered the tricky State 6 and State 7, those given the aligned operators were not tempted to try a different configuration. They remained focused on the key part of the problem, and found the way to the goal with fewer moves.

Solving problems in the real world can benefit from similar design choices. A control or display panel should incorporate as many real restrictions as possible. The interactive elements should also be designed to match how the rules function. Having more of the problem embedded in the interface reduces the effort required to adhere to the rules and frees more resources for finding the solution. The incorrect perception of equivalent configurations as new, distinct states may be automatic. Some extra effort to counteract this perception would be beneficial.

Learning the terminology in a new field is clearly important, and is required for mastery. However, the equivalence that is easily seen by experts is cannot be assumed to understood by those entering the field. Topics must be prioritized when learning a new field. Understanding the equivalency or understanding the underlying logic are easier if tackled separately. Learning becomes much more difficult when they are presented simultaneously. Taking steps to limit the generation of unnecessarily complex representations would allow people to find solutions more quickly.

REFERENCES

Anderson, J. R., & Douglass, S. (2001). Tower of hanoi: Evidence for the cost of goal retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1331-1346.

Atwood, M., & Polson, P. G. (1976). A process model for water jug problems. *Cognitive Psychology*, *8*(2), 191–216.

Gentner, D. (1983). Structure mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155 - 170.

Greeno, J. G. (1974). Hobbits and orcs: Acquisition of a sequential concept. *Cognitive Psychology*, *6*(2), 270 - 292.

Hayes, J. R., & Simon, H. A. (1974). Understanding written problem instructions. In *Knowledge and cognition.* Lawrence Erlbaum.

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*(4), 412 - 440.

Knowles, M. E., & 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*(4), 670 - 682.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? evidence from tower of hanoi. *Cognitive Psychology*, *17*(2), 248 - 294.

Kotovsky, K., & Simon, H. A. (1990). What makes some problems really hard: Explorations in the problem space of difficulty. *Cognitive Psychology*, *22*(2), 143 - 183.

Luchins, A. S. (1942). Mechanization in problem solvingthe effect of einstellung. *Psychological Monographs*, *546*, 95.

Metcalfe, J. (1986). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(2), 288-294.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Prentice-Hall.

Ramsey, F., & Schafer, D. (1997). The statistical sleuth: A course in methods of data analysis. In (1st ed., p. 632 - 638). Duxbury Press.

Reed, S. K., Ernst, G. W., & Banerji, R. (1974). The role of analogy in transfer between similar problem states. *Cognitive Psychology*, *6*(3), 436 - 450.

Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(5), 484 - 494.

Simon, H. A., & Hayes, J. R. (1976). The understanding process: Problem isomorphs. *Cognitive Psychology*, *8*(2), 165 - 190.

Simon, H. A., & Reed, S. K. (1976). Modeling strategy shifts in a problem-solving task. *Cognitive Psychology*, *8*(1), 86 - 97.

Thomas, J. C. (1974). An analysis of behavior in the hobbits-orcs problem. *Cognitive Psychology*, *6*(2), 257 - 269.

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

APPENDIX A

TABLES

Table A1

Problem Rules

Experiments 1 and 3: Grance and Zurkistan

1. Neither capital was ever destroyed.

2. Whenever two of Grance's cities were wiped out, exactly two of Zurkistan's cities were wiped out.

3. Whenever three of Grance's cities were standing, exactly three of Zurkistan's cities were standing.

4. One or two cities were destroyed or rebuilt each season.

Experiment 2: Hobbits and Orcs

1. One traveler must be in the boat to move the boat across the river.

2. The boat can carry a maximum of two travelers.

3. At no time can the orcs outnumber the hobbits on either side of the river.

Table A2

Experiment 1 Move Counts

Selection Order	Legal Moves	Rule Violations
Open	35.95 (34.44)	24.13 (12.44)
Fixed	21.51 (16.30)	17.23 (11.40)

Note. Values given are untransformed. Normal distribution standard deviations are in parentheses.

Table A3

Experiment 1 Move Reversals

Selection Order	Exact	Total
Open	12.00 (13.11)	22.32 (23.37)
Fixed	11.23 (10.05)	–

Note. Total reversals include equivalent and exact reversals. Values given are untransformed. Normal distribution standard deviations are given in parentheses.

Table A4

Experiment 2 Non-Solver Rates

Condition	Solved	Ran out of time	Percentage
Open-Identical	40	2	4.8%
Fixed-Identical	36	1	2.7%
Open-Distinct	33	3	8.3%
Fixed-Distinct	30	2	6.3%

Table A5

Experiment 2 Mean Move Counts

Condition	Legal Moves	Rule Violations
Open-Identical	27.95 (15.72)	15.35 (10.21)
Fixed-Identical	24.22 (9.70)	14.58 (11.18)
Open-Distinct	27.36 (16.83)	14.64 (12.58)
Fixed-Distinct	25.23 (13.90)	15.77 (16.30)

Note. Values given are untransformed. Normal distribution standard deviations are in parentheses.

Table A6

Experiment 2 Mean Reversal Counts

Condition	Exact	Total
Open-Identical	7.47 (4.93)	10.85 (7.66)
Fixed-Identical	8.25 (4.22)	–
Open-Distinct	6.76 (4.92)	11.03 (8.10)
Fixed-Distinct	9.20 (8.36)	–

Note. Total reversals include equivalent and exact reversals. Equivalent reversals were not possible in the Fixed conditions. Values given are untransformed. Normal distribution standard deviations are in parentheses.

Table A7

Experiment 3 Non-Solver Rates

Condition	Solved	Ran out of time	Percentage
Move - Open	21	4	16.0%
Move- Fixed	15	7	31.8%
Change - Open	16	2	11.1%
Change - Fixed	18	2	10.0%

Table A8

Experiment 3 Mean Move Counts

Condition	n	Legal Moves	Rule Violations
Move - Open	21	21.67 (12.78)	22.38 (13.48)
Move- Fixed	14	22.21 (14.46)	23.29 (10.80)
Change - Open	16	25.31 (21.42)	20.56 (13.59)
Change - Fixed	15	21.00 (13.09)	22.33 (11.74)

Note. Values given are untransformed. Normal distribution standard deviations are in parentheses.

Table A9

Experiment 3 Reversal Counts

Condition	Exact	Total
Move - Open	5.52 (3.28)	11.81 (9.51)
Move - Fixed	9.50 (6.54)	–
Change - Open	7.19 (6.36)	13.38 (13.04)
Change - Fixed	10.47 (7.09)	–

Note. Total reversals include equivalent and exact reversals. Equivalent reversals were not possible in the Fixed conditions. Values given are untransformed. Normal distribution standard deviations are in parentheses.

APPENDIX B

FIGURES

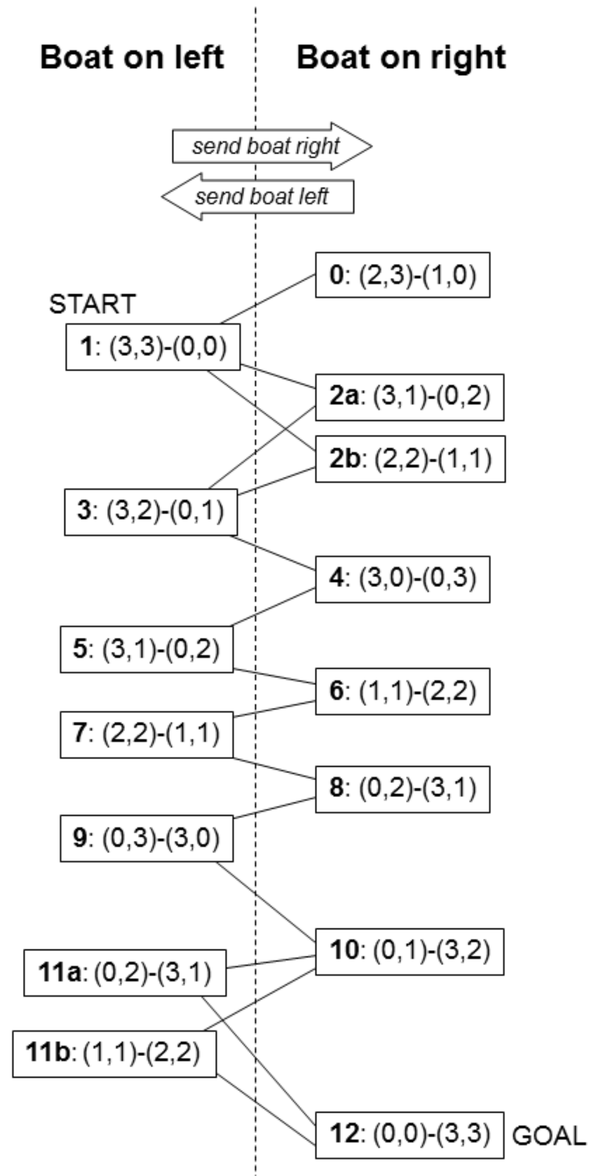

Figure B1. Hobbits and Orcs problem space. Rectangles represent states. Bold numbers are the state names. The hyphenated numbers are the number of hobbits and orcs, respectively, on each river bank.

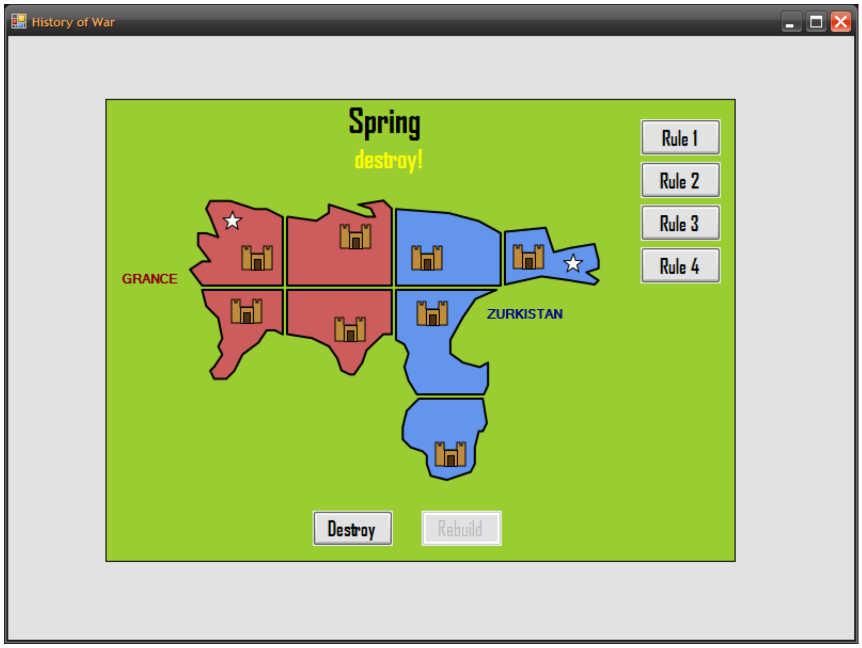

Figure B2. Grance and Zurkistan Open condition.

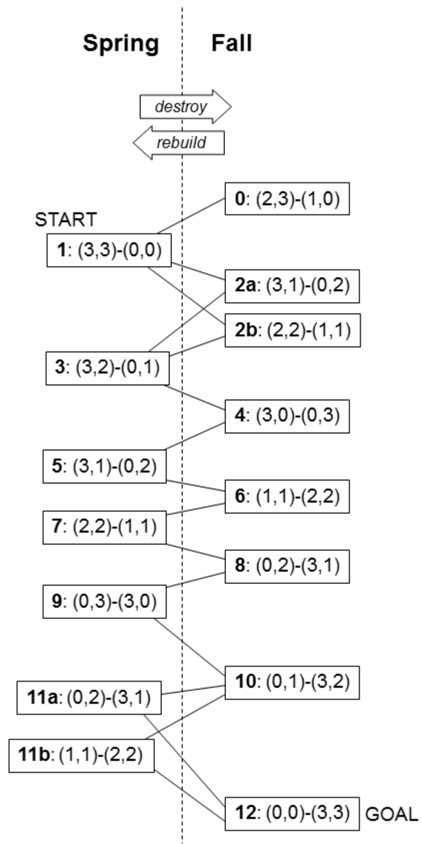

Figure B3. Grance and Zurkistan problem space. Rectangles represent states. Bold numbers are the state names. The hyphenated numbers are the counts of cities in Grance and Zurkistan that are intact, followed by the numbers of cities in Grance and Zurkistan that were destroyed.

Figure B4. Grance and Zurkistan Fixed condition with two highlighted cities.

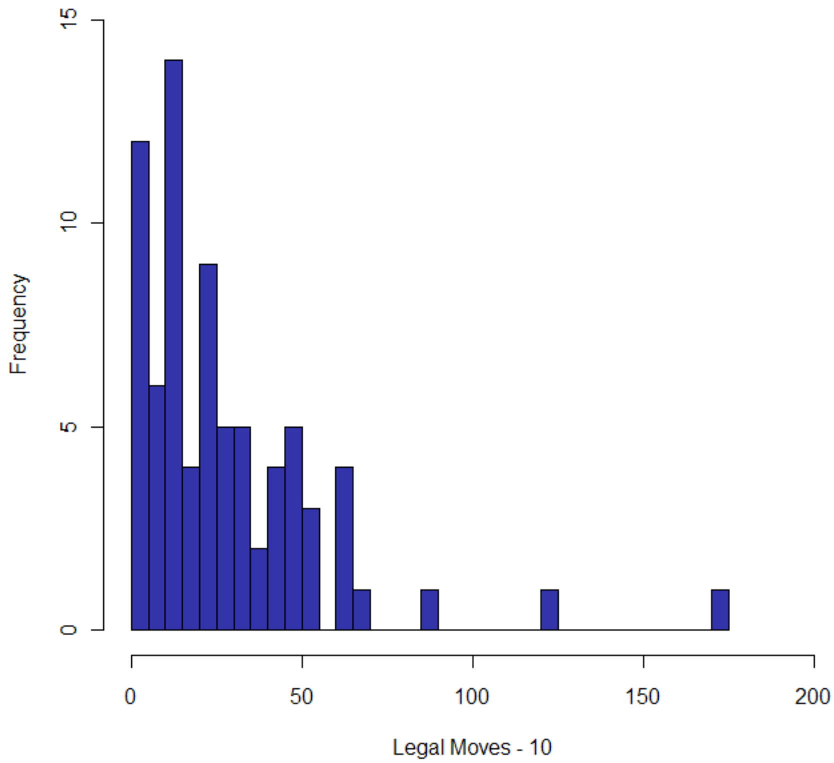

Figure B5. Experiment 1 legal moves histogram. One less than the minimum number of required moves was subtracted from the legal move counts to shift the distribution down to near zero.

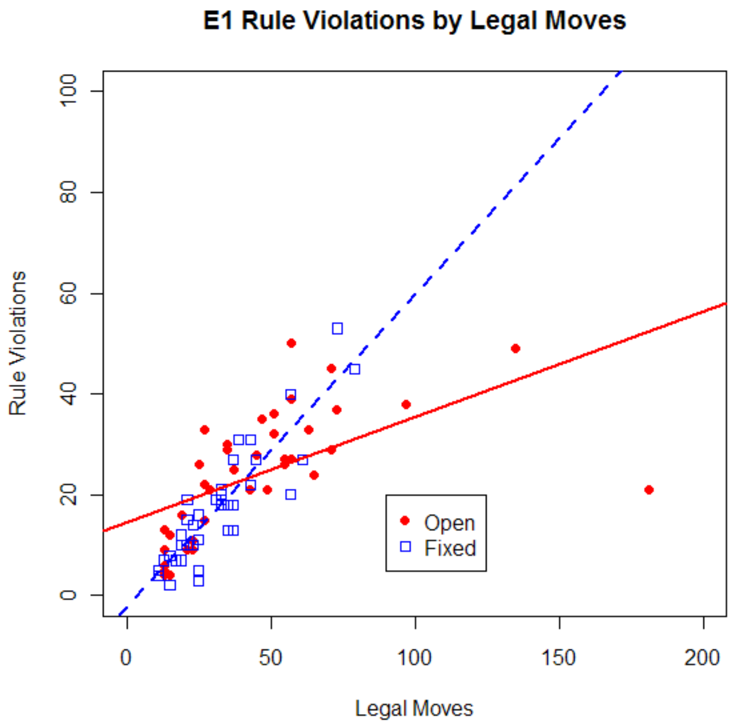

Figure B6. Experiment 1 linear regression of rule violations on legal moves.

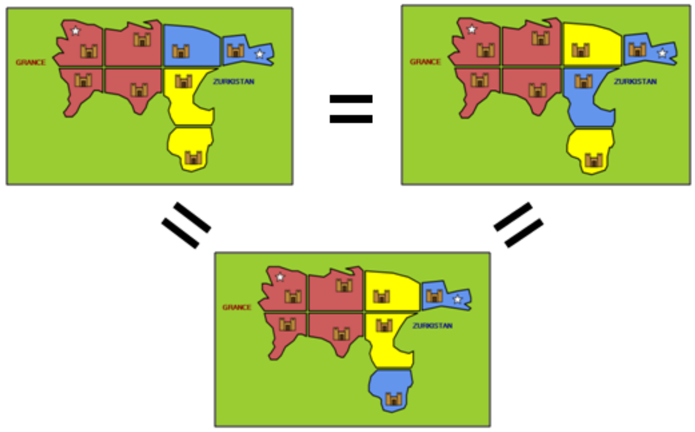

Figure B7. Example of equivalent states.

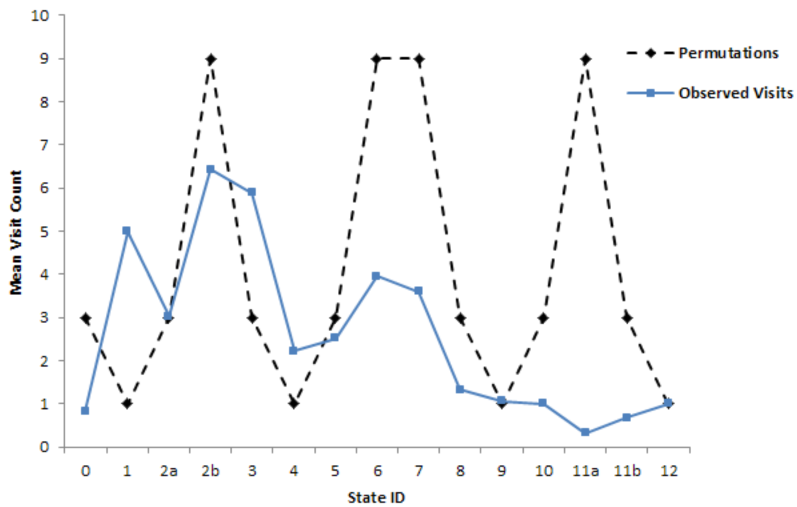

Figure B8. Experiment 1 observed and potential visit counts per state. The solution path progress from left to right, start to goal, along the x-axis. The mean visit counts per state is shown on the y-axis. The number of possible combinations of destroyed cities for each equivalent state are shown by the dashed line.

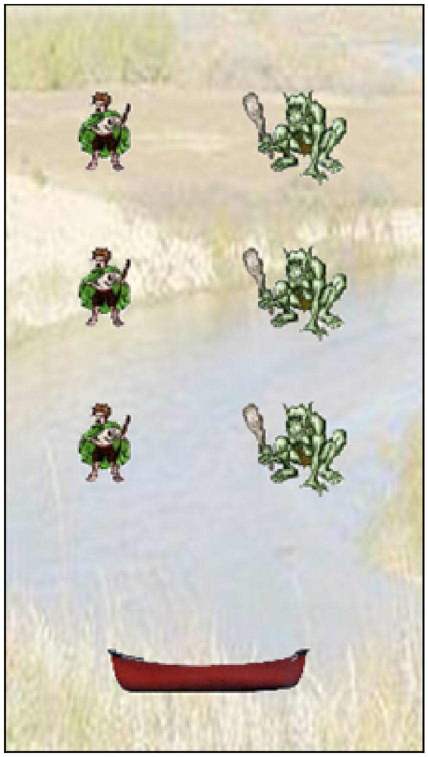

Figure B9. Experiment 2 character appearances. Identical hobbits and orcs are shown on the left. Distinct hobbits and orcs are shown on the right.

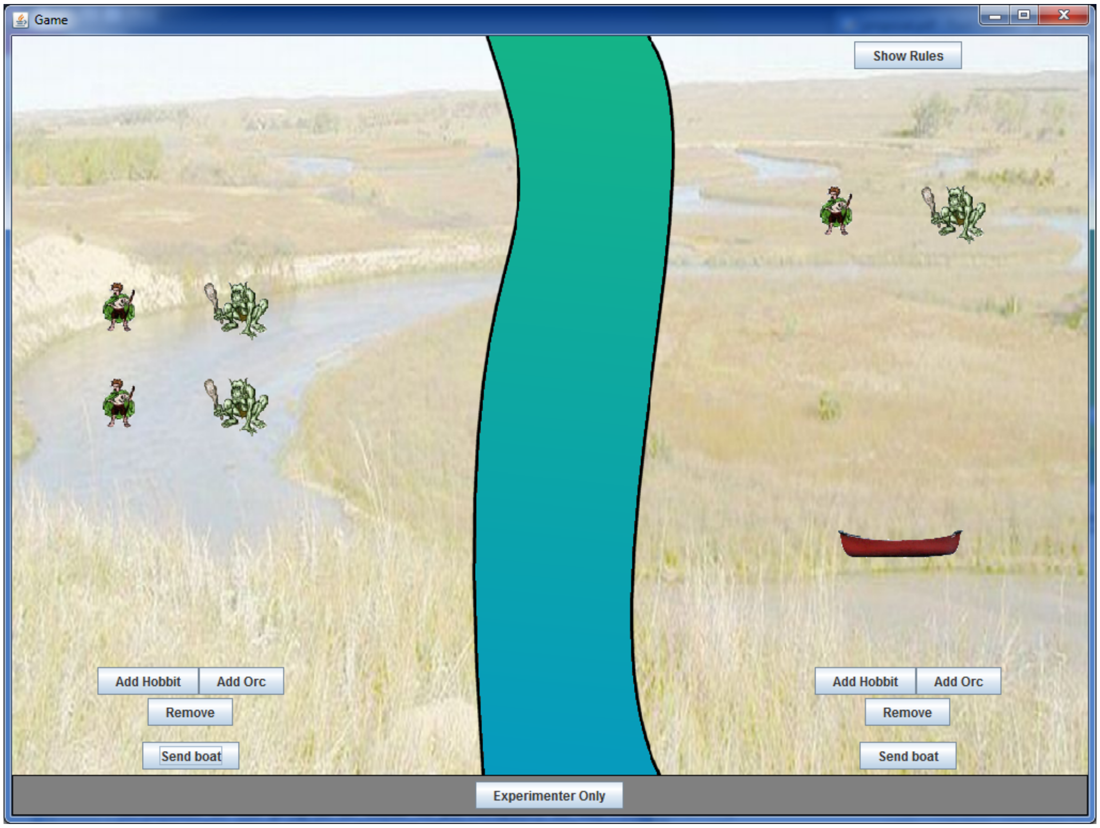

Figure B10. Fixed version of Hobbits and Orcs.

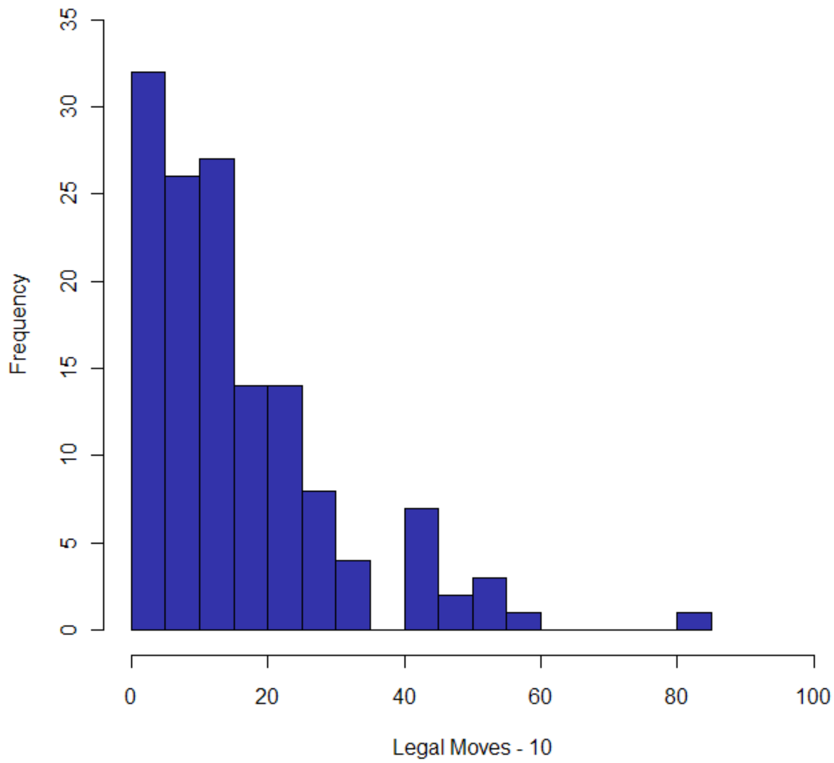

Figure B11. Experiment 2 legal moves histogram. One less than the minimum number of required moves was subtracted from the legal move counts to shift the distribution down to near zero.

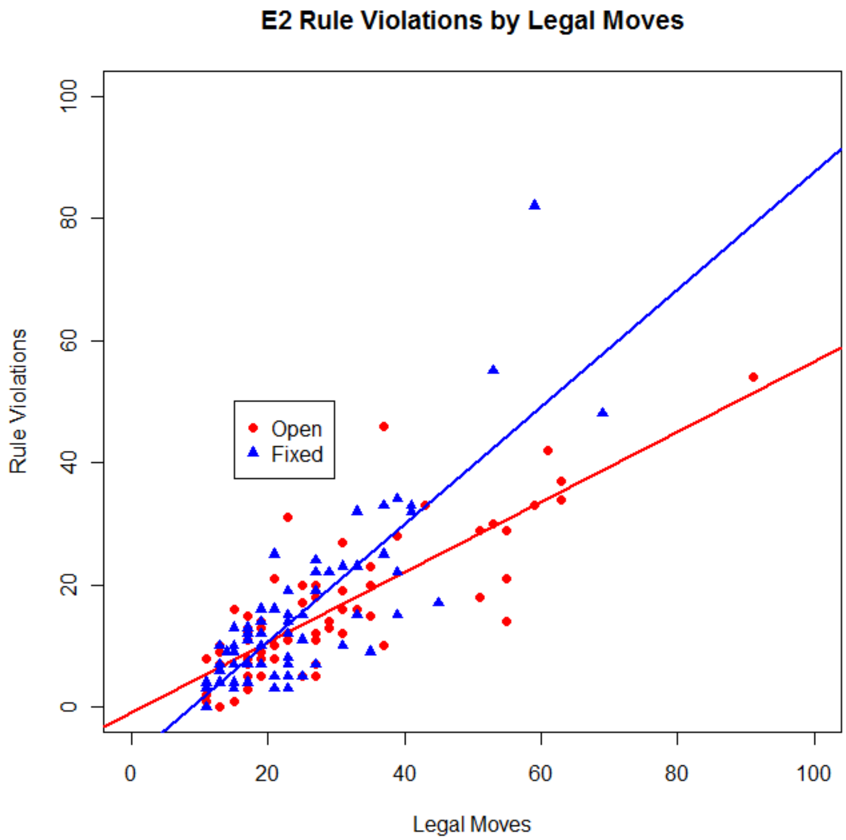

Figure B12. Experiment 2 open and fixed linear regressions.

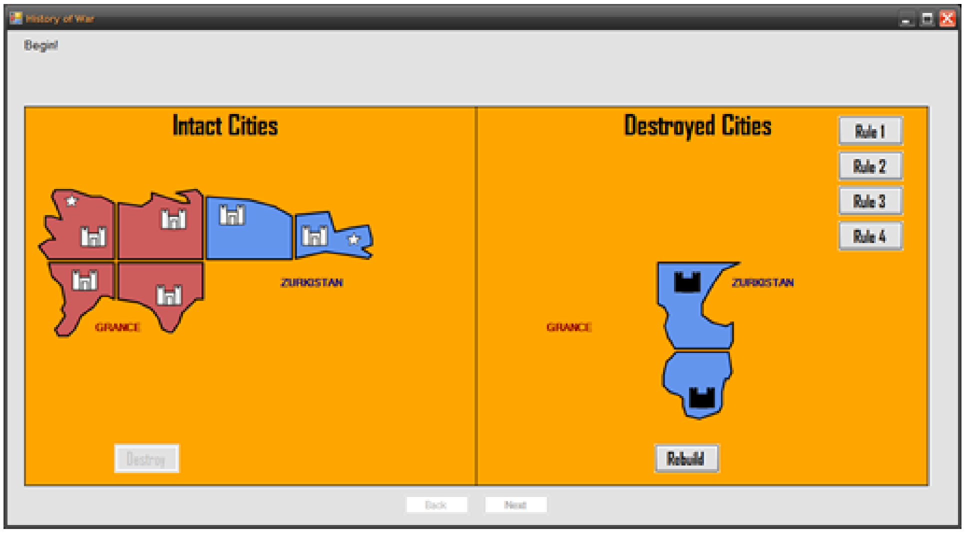

Figure B13. Move version of the Grance and Zurkistan problem.

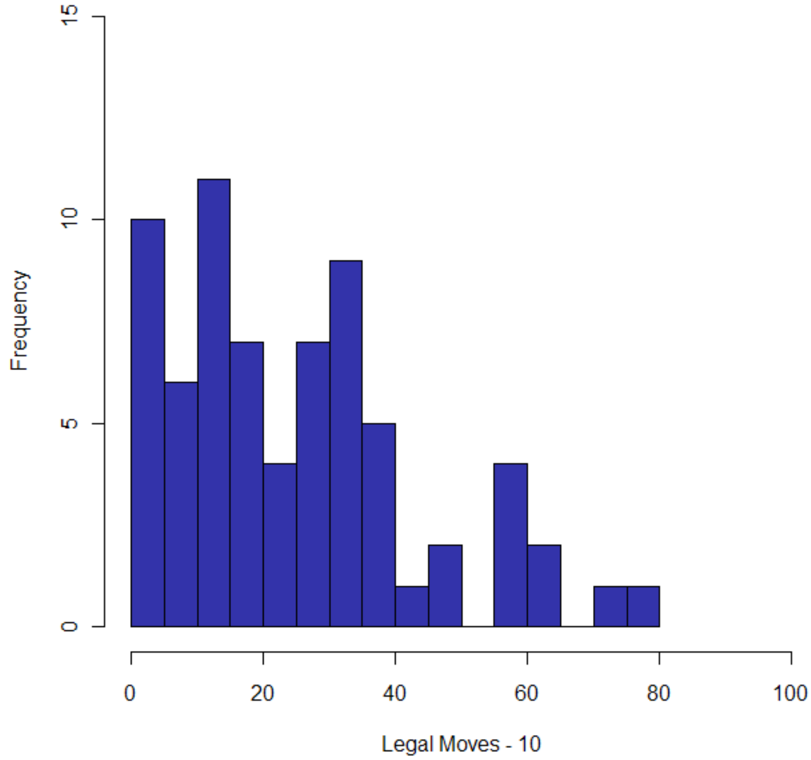

Figure B14. Experiment 3 legal moves histogram. One less than the minimum number of required moves was subtracted from the legal move counts to shift the distribution down to near zero.

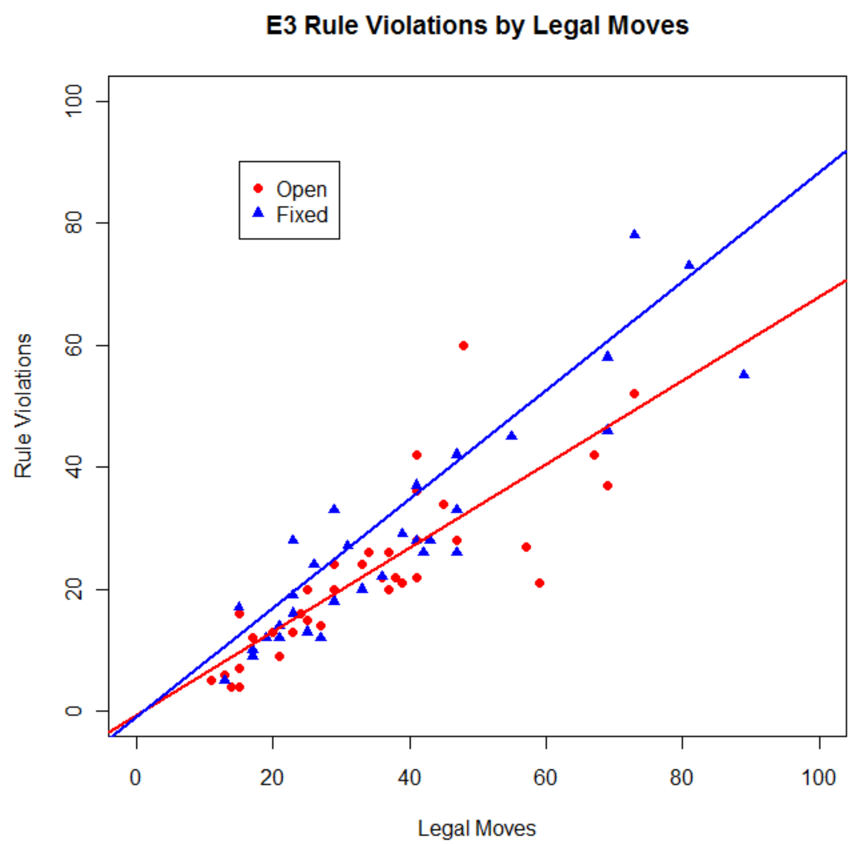

Figure B15. Experiment 3 Open and Fixed linear regressions.

APPENDIX C

EXPERIMENT 2 TRAINING SCRIPTS

Open-Identical. The goal of this puzzle is to move all of the travelers safely across the river. The hobbits are the small, friendly travelers lined up on the left. The orcs are mean and aggressive, and are to the right of the hobbits. Practice by putting an orc in the boat by clicking on one of the orcs. Now move that orc across by clicking the "Send boat" button. Click on the orc on the right bank to put him in the boat and move him back to the left. This time click on a hobbit and an orc to put two travelers in the boat. Send them over to the right. Finally, add the orc on the right to the boat and send it back to the left. Note the orcs now outnumber the hobbits on the left bank, breaking the third rule.

*Fixed-Identical.*The goal of this puzzle is to move all of the travelers safely across the river. The hobbits are the small, friendly travelers lined up on the left. The orcs are mean and aggressive, and are to the right of the hobbits. Practice by putting an orc in the boat by clicking the "Add Orc" button. Now move that orc across by clicking the "Send boat" button. Click on the "Add Orc" button to put the orc on the right bank in the boat and move him back to the left. This time add one hobbit and one orc to the boat. Send them over to the right. Finally, add the orc on the right to the boat and send it back to the left. Note the orcs now outnumber the hobbits on the left, breaking the third rule.

*Open-Distinct.*The goal of this puzzle is to move all of the travelers safely across the river. The hobbits are the small, friendly travelers lined up on the left. The orcs are mean and aggressive, and are to the right of the hobbits. Practice by putting Zob in the boat by clicking on Zob. Now move Zob across by clicking the "Send boat" button. Click on Zob to put him in the boat and move him back to the left. This time click on Rak and Joe to put them in the boat. Send them both over to the right. Finally, add Rak to the boat and send it back to the left. Note Zob, Rak, and Lex now outnumber Sam and Pip on the left, breaking the third rule.

*Fixed-Distinct.*The goal of this puzzle is to move all of the travelers safely across the river. The hobbits are the small, friendly travelers lined up on the left. The orcs are mean and aggressive, and are to the right of the hobbits. Practice by putting Zob in the boat by clicking the "Add Orc" button. Now move Zob across by clicking the "Send boat" button. Click on the "Add Orc" button to put Zob in the boat and move him back to the left. This time click on the "Add Hobbit" and "Add Orc" buttons to put Zob and Sam in the boat. Send them both over to the right. Finally, add Zob to the boat and send it back to the left. Note Zob, Rak, and Lex now outnumber Pip and Joe on the left, breaking the third rule.

APPENDIX D

GRANCE AND ZURKISTAN TRAINING SCREENS

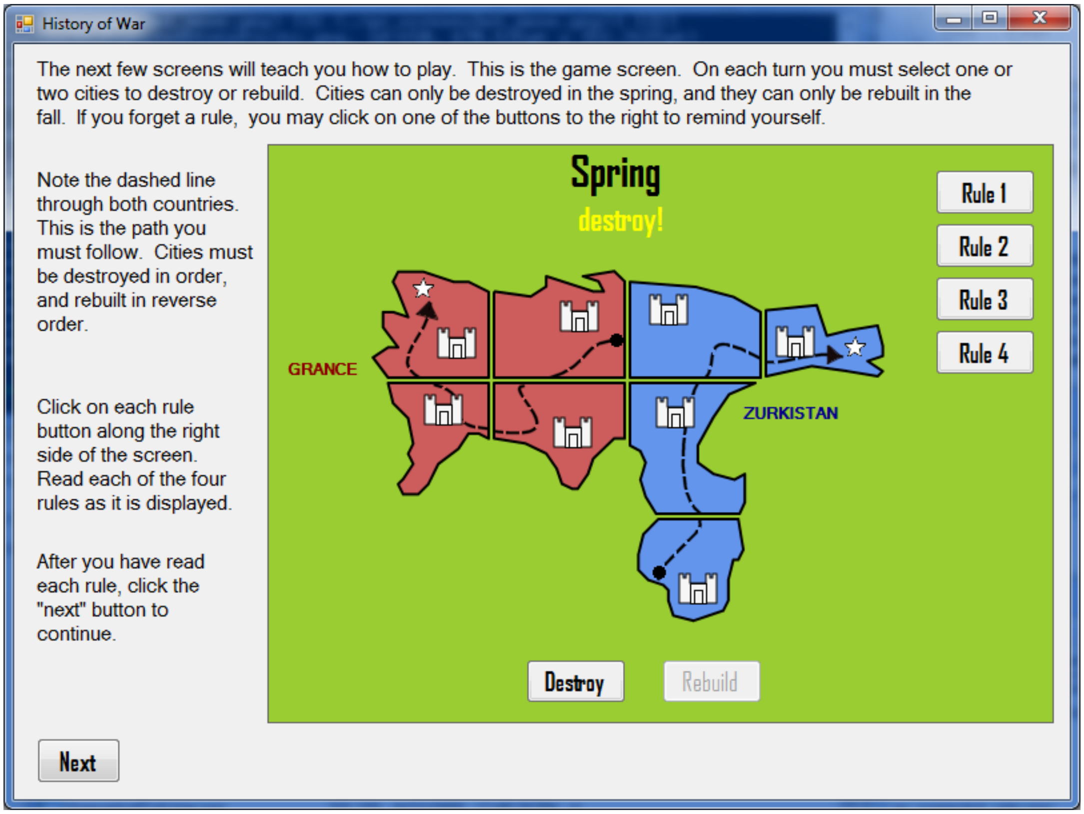

Figure D1. Program introduction screen.

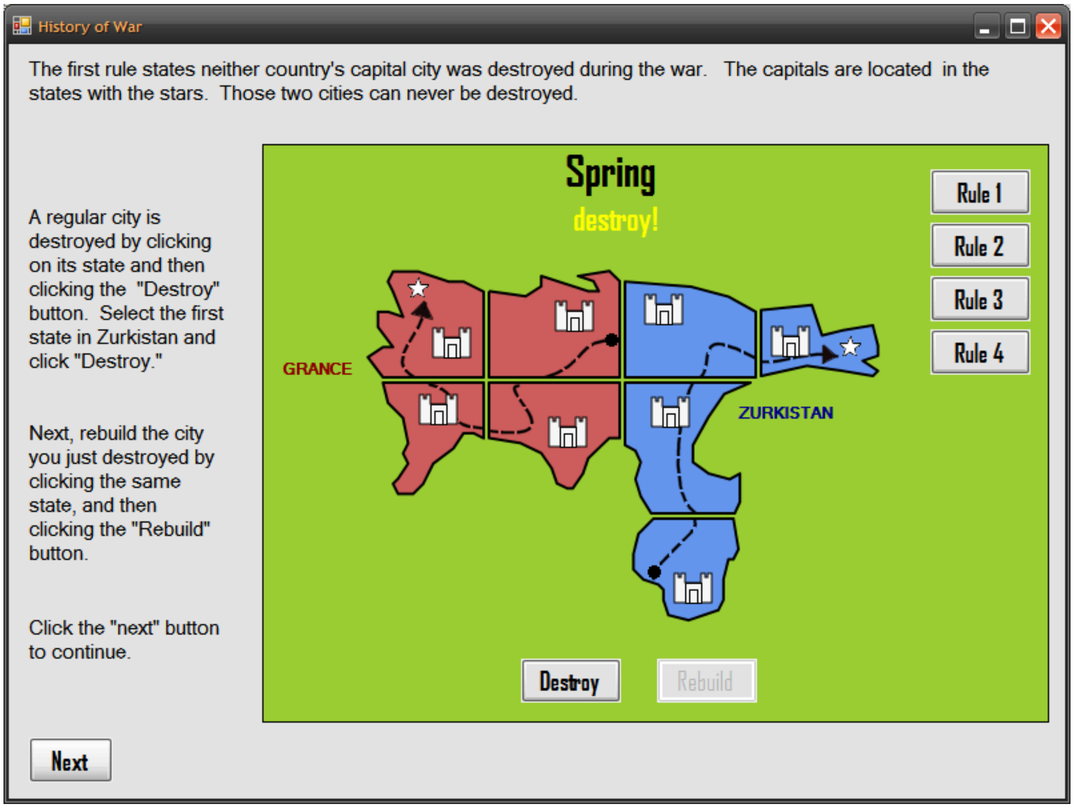

Figure D2. Training screen for Rule 1.

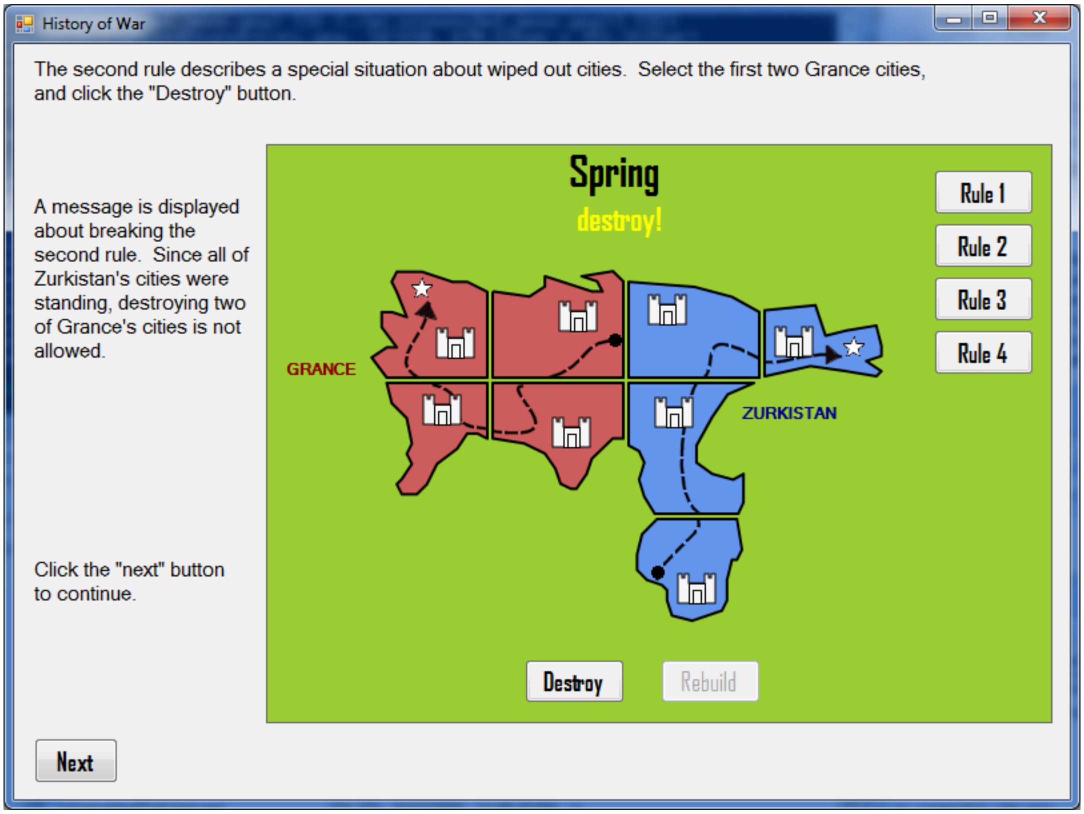

Figure D3. Training screen for Rule 2.

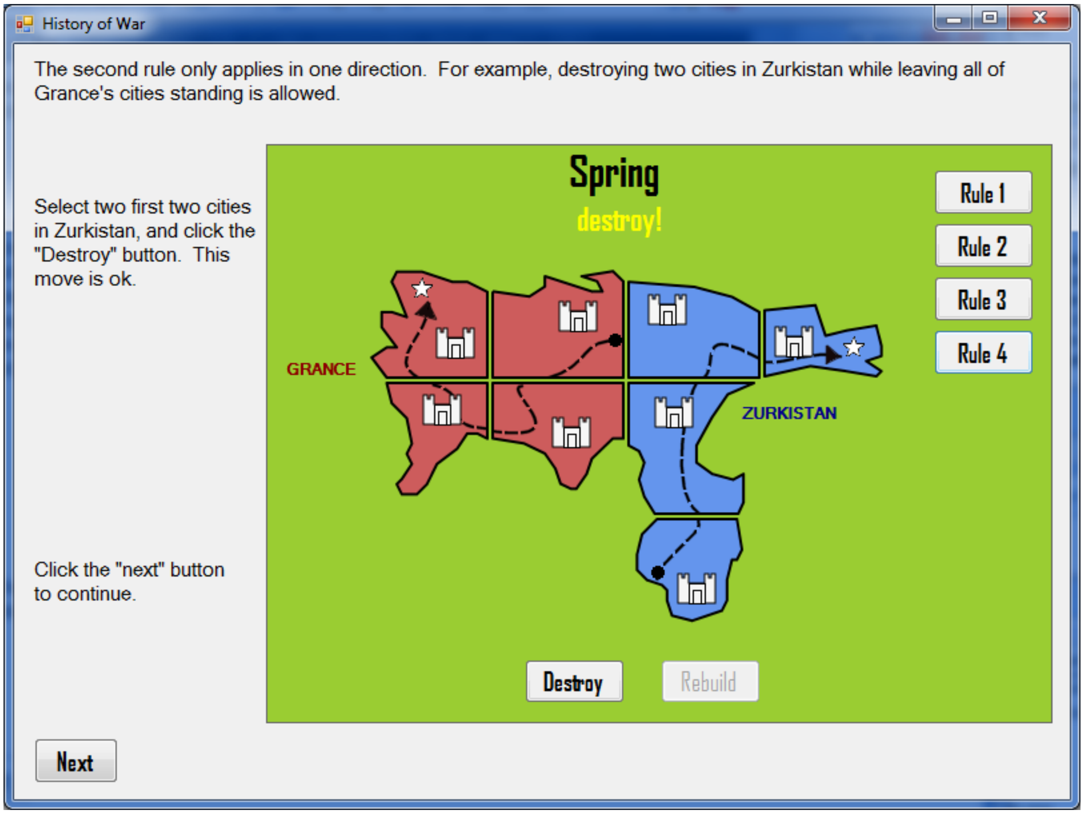

Figure D4. Second example for Rule 2.

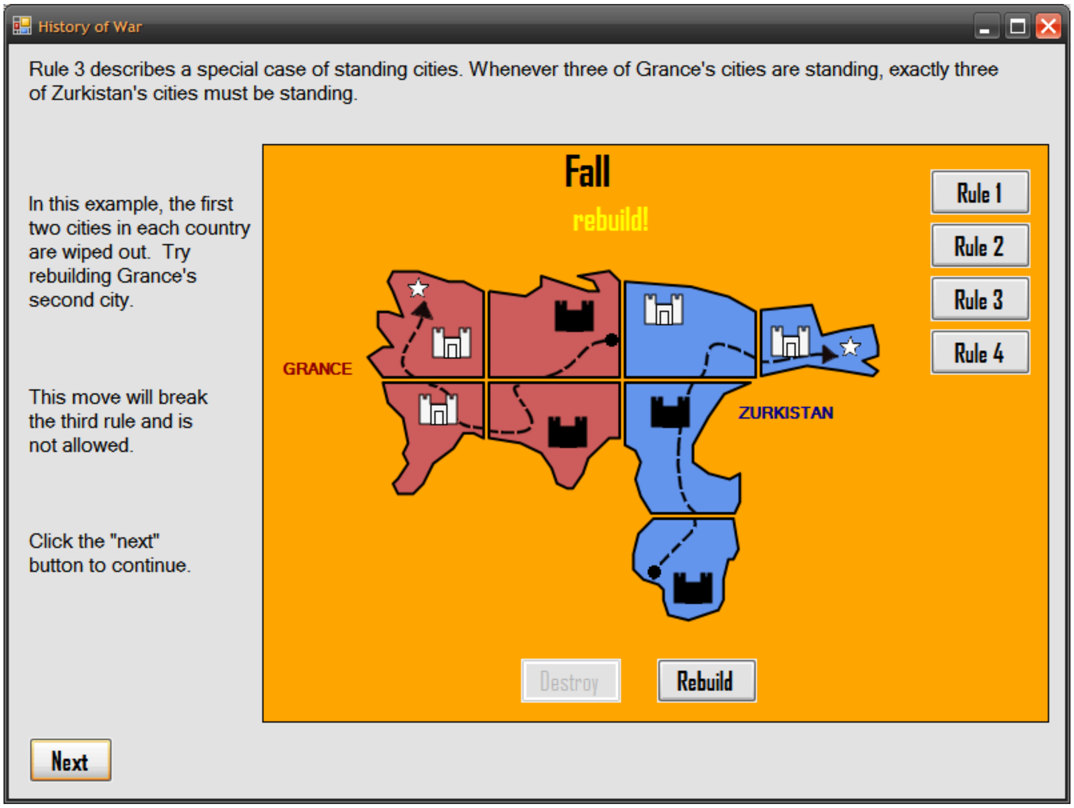

Figure D5. Training screen for Rule 3.

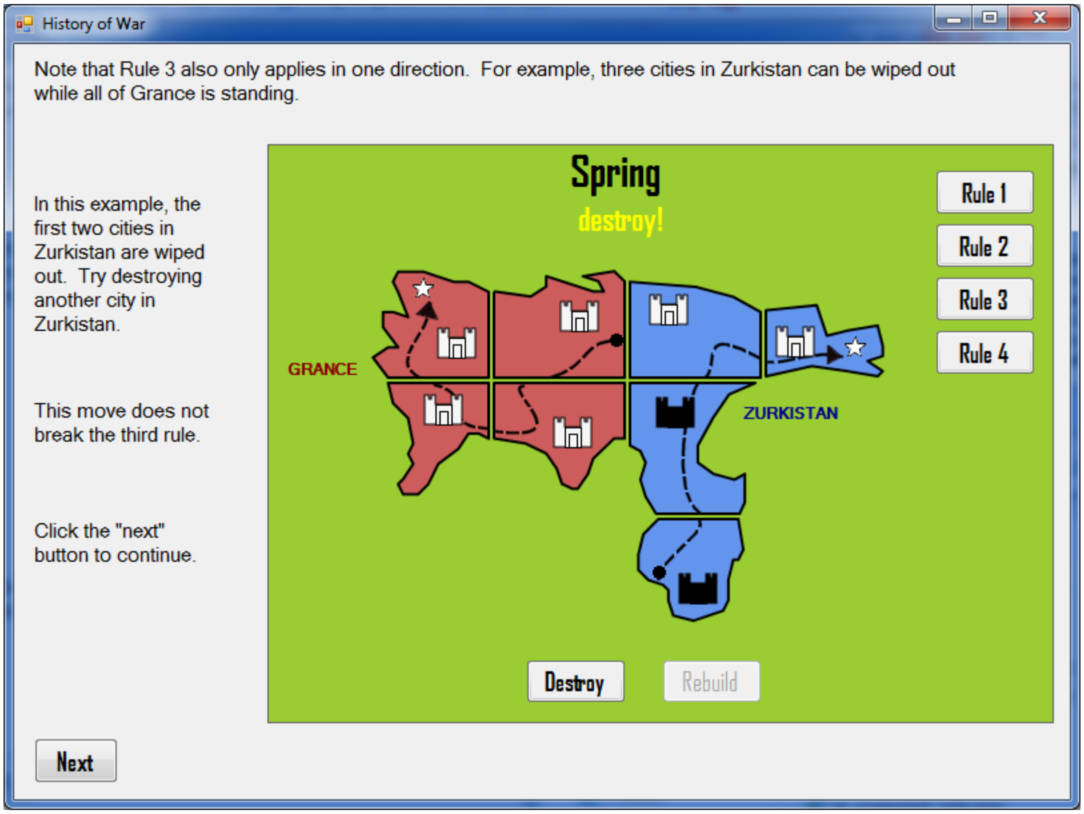

Figure D6. Second example for Rule 3.

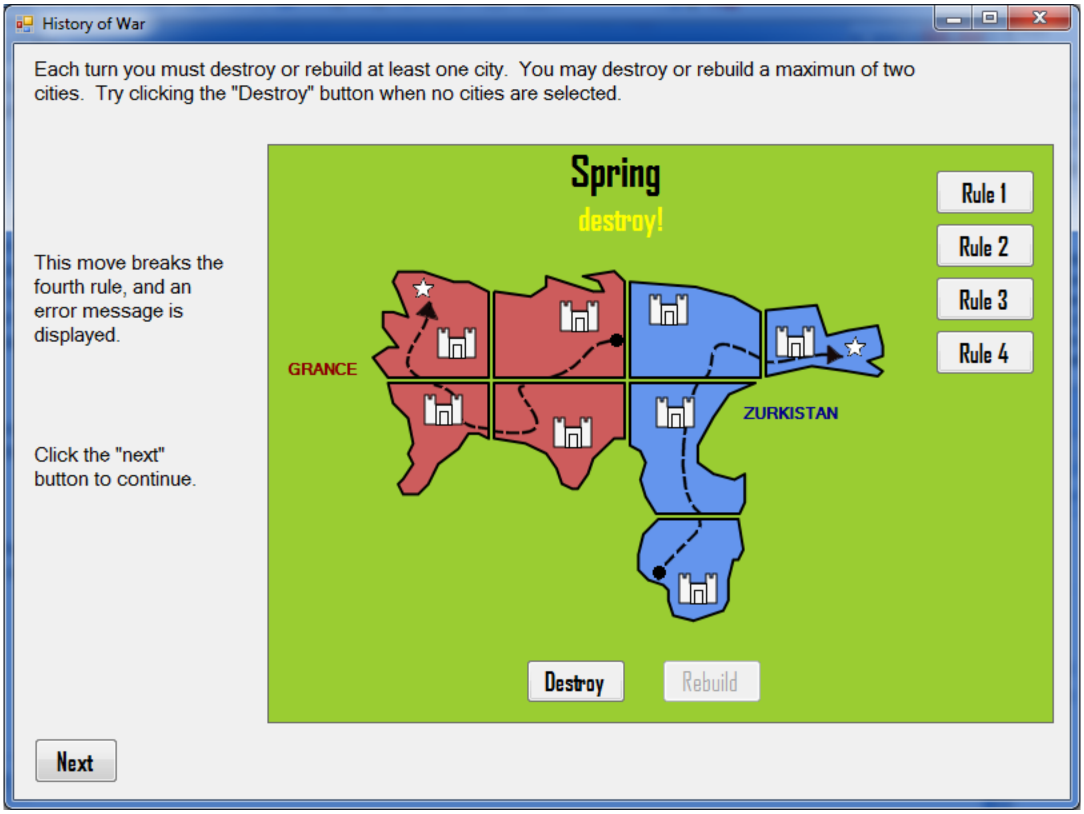

Figure D7. Training screen for Rule 4.

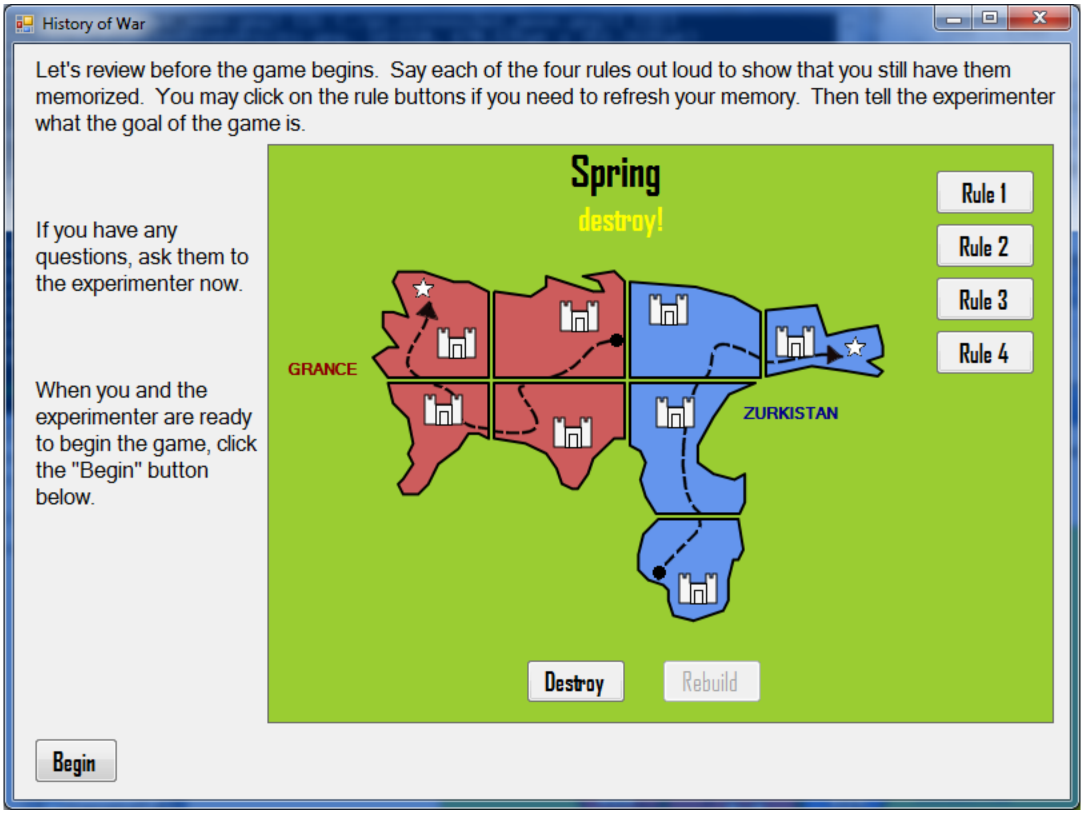

Figure D8. Training summary screen.