INFORMATION TO USERS

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

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

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

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

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

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

TRANSFER BETWEEN DIFFERENT CONTEXTS:
COMPARING INTERACTIVE AND NON-INTERACTIVE TRAINING.

BY

BERNARD H CHEN

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

MASTER OF ARTS

APPROVED, THESIS COMMITTEE:

David Lane, Director
Associate Professor of Psychology

Miguel Quinones
Associate Professor of Psychology

John Brelsford
Professor of Psychology

Houston, Texas
February, 1999
ABSTRACT

Transfer Between Different Contexts:
Comparing Interactive and Non-Interactive Training

by

BERNARD H CHEN

This study attempted to replicate the findings of one such case of far transfer by Fong and Nisbett (1986) in order to support the position that far transfer can be reliably achieved. The requirements for transfer, in terms of superficially and structurally similar elements, are stated, leading to a hypothesis that interactive training will lead to higher rates of transfer. Ninety-four undergraduate students from Rice University were assigned to three training groups, a non-interactive Expository group, an interactive Interactive-Simulation group, and an observational, Passive-Simulation group. A one-week delay was also used. Transfer of training was found, although Interactive training did not lead to higher scores than Expository training. Several factors that influenced the likelihood of transfer are identified and discussed. Additional considerations are also made regarding features of training that emphasize generalizability.
This thesis is dedicated to:

Helen and P.J.

Thanks mom and dad.
# TABLE OF CONTENTS

Introduction .................................................................................................................. 1

Theories of Transfer ..................................................................................................... 2

Interactive Training .................................................................................................... 18

Method ....................................................................................................................... 26

Participants .................................................................................................................. 26

Procedure ..................................................................................................................... 26

Materials ...................................................................................................................... 37

Results ......................................................................................................................... 33

Factors of the Law of Large Numbers ....................................................................... 38

Overall Transfer ......................................................................................................... 44

The Human Performance factor ................................................................................. 47

The Small Sample factor ........................................................................................... 50

The Large Sample factor ........................................................................................... 52

Discussion ................................................................................................................... 55

References ................................................................................................................... 65

Appendix A: Law of large numbers examples ............................................................ 70

Appendix A.1: Abstract example of the law of large numbers .................................... 70

Appendix A.2: Concrete example of the law of large numbers .................................... 71

Appendix B: Eight Transfer Questions ........................................................................ 74

Appendix B.1: Tryouts ................................................................................................. 74

Appendix B.2: Slots ..................................................................................................... 75
Appendix B.3: Cars .................................................................................................. 76
Appendix B.4: Darts ................................................................................................. 77
Appendix B.5: Babies ............................................................................................... 78
Appendix B.6: Auditions ......................................................................................... 79
Appendix B.7: Admissions ....................................................................................... 80
Appendix B.8: Stocks .............................................................................................. 71
Appendix C: Partial correlation of transfer questions, controlling for training type. .... 82
Appendix D: Example three for the Expository group. ............................................ 83
LIST OF TABLES

Table 1: Advantages of interactive teaching materials. 19

Table 2: The transfer questions did not load cleanly onto the three factors proposed by Fong and Nisbett (1986). 39

Table 3: Factor loadings of the eight transfer questions. Three factors provided the best fit. 40

Table 4: A partial correlation matrix showing that the pattern of correlations supports the use of three factors. 40

Table 5: Questions comprising the three factors identified by the factor analysis correlated more within each factor than with questions outside of the factors 42
LIST OF FIGURES

Figure 1: Transfer scores from Fong and Nisbett (1986). 15
Figure 2: Advice too abstract to apply. 17
Figure 3: Phase one of the card simulator. The subject can manipulate several variables. 29
Figure 4: Phase two of the card simulator. The subject must determine if the deck is Honest or Cheaters. 31
Figure 5: Percentage of correct responses in the first and last five rounds of the card simulator. 34
Figure 6: Number of hands dealt by the subject in the first and last five rounds of the card simulator. 35
Figure 7: Results of simulated strategies for the card simulator. 36
Figure 8: Average scores for each of the Training Groups and the control group. 44
Figure 9: Average scores for each group, divided by Cognitive Ability. 45
Figure 10: Average scores for each group, divided by Delay 46
Figure 11: Average scores for each group on the Human Performance factor. 47
Figure 12: Average scores for each group on the Human Performance factor, divided by Cognitive Ability. 48
Figure 13: Average scores for each group on the Small Samples factor. 51
Figure 14: Average scores for each group on the Large Samples factor. 53
Figure 15: Average scores for all three LLN factors. 54
Introduction

The goal of training and instruction is to impart new knowledge and skills to an individual. In addition, it is important that this knowledge is applied to similar problems and situations outside of the training context as well as to problems with dissimilar, yet related contexts. The bridging of familiar and unfamiliar problems is the focus of transfer research. Numerous studies have examined the transfer problem, where transfer has been defined as the application of learned procedures in novel scenarios. The results have been mixed. Although attempts have been made to understand the process of transfer, a single definition has not arisen.

One aspect that may affect the degree of transfer is the extent to which training is interactive. A common approach in transfer research is to provide training in a non-interactive format in which the student plays the role of a passive receiver of information. Instead, there are reasons to believe that expanding the role of the student to that of an active participant can improve learning and increase the likelihood of transfer (Rivera-Perez, 1996).

The purpose of this study is to examine the effect of interactive training on subjects' abilities to transfer a statistical principle, the law of large numbers (LLN). First, a review of the theories of transfer will be presented. Second, findings regarding transfer will be discussed. Third, the dimensions of interactive training will be examined, and fourth, a hypothesis regarding the effects of interactive training on transfer will be presented.
Theories of Transfer

Transfer is the exercising of learned concepts in subsequent situations. Although two situations can differ in terms of context, separation in time, or other details, most new situations have something in common with a past experience. Consequently, dealing with new situations infrequently requires an entirely unique solution. Instead, variations of solutions used in the past can be reused.

There is a popular two-part format for studying transfer. Typically, a training group is taught a skill during a Training Phase using a particular task or a set of tasks. In a later Transfer Phase, subjects’ understanding of the original skill is tested using tasks that differ in degrees from the ones initially used. As a comparison, a control group of subjects is presented with only the tasks from the transfer phase without prior training. There are three possible results from such a format.

Occasionally, aspects of training cause the training group to perform worse on the transfer task, a condition called negative transfer. For instance, Hammond (1991) uses the example of training on a typewriter. Typewriter-trained typists will reach for the carriage return on a computer keyboard even though the response is unnecessary and incorrect. Compared to a group trained solely on keyboards, they are at a relative disadvantage. However, incidents of negative transfer commonly exist in a background of positive transfer, the second possible result from transfer tasks. Even though the typewriter-trained group performs worse than the keyboard-trained group, their typing skill is still greater than an untrained control group.
The last possible outcome from the transfer task is that there is no difference between the performance of the training group and the control group. In these cases, the differences between the training and transfer tasks are too great for subjects to reconcile.

Research on transfer exists in a region that is bordered by transfer between identical situations on one side, and the lack of transfer between dichotomous situations on the other. Naturally, this extends itself to a definition of transfer based on the similarity between tasks.

**Near and Far transfer**

The likelihood of transfer can be described using a metaphor of distance. Transfer between highly related problems, near transfer, has been the easiest to produce, while its opposite, far transfer, is more eagerly sought and harder to document. Although near and far exist on the same continuum, they do illustrate two basic categories of advantages and disadvantages. If instructors could depend on near transfer, they could improve the quality of performance of their students, but only for problems similar to the ones used during training. In contrast, far transfer is alluring because, if it could be reliably depended upon, lessons learned from a few examples would be generalized to a wide range of scenarios, thus amplifying the effects of training.

Near transfer is commonly instantiated by presenting a transfer task in the same context as the training task. If a training problem contained a surgeon and a tumor, then a near transfer problem would also contain a surgeon and a tumor. The repetition of the context and objects in the transfer problem cue the recall of the original training problem, reminding the subject of a set of information that may apply to the current problem. In
fact, the major benefit from near transfer is that the new problem is somehow associated with old, applicable information. Historically, subjects are far more likely to repeat past solutions to a new problem when they realize that it is similar to an old problem (e.g., Gott, Hall, Pokorny, Dibble, and Glaser, 1993).

Far transfer exists on the opposite end of the transfer continuum. As the similarities between two problems decrease, the likelihood that information about one will be applied to the other also decreases. Most commonly, the context of training and transfer problems are differ. If subjects were once again trained a principle set in the context of a surgeon and a tumor, a far transfer problem could be set in the context of a general and a military target (Gick and Holyoak, 1983.) Greeno, Smith, and Moore (1993, in D) hypothesize that changes in context are frequently enough to prevent the transfer of learning because people learn how to act according to situations. Experience teaches people the affordances and invariants of a particular setting and it is difficult for people to overcome this context-dependency.

Very few examples of far transfer experiments are present in the literature, opening up the possibility that far transfer is an anomaly. Detterman (1993) reviewed the far transfer literature only to conclude that the few cases of documented far transfer could be attributed to demand characteristics or non-subtle interventions by the experimenter. Even so, successful cases of far transfer share enough consistencies to provide the basis for theories on the mechanism of transfer and to support a belief that far transfer can be consistently achieved.

**Superficial and Structural Features**
Early learning theorists adopted the view that the mind was a muscle that improved with use and that general learning could occur without specifying constraints for transfer. This perspective, however, met with little success. Thorndike and Woodworth (1901) proposed a new perspective on general learning by stating that transfer was dependent on the presence of identical elements between tasks. This theory was an improvement, but was still more descriptive than predictive, because the only way to prove that elements in two problems were identical was if transfer occurred.

Recent researchers have suggested that similarities between certain types of features determine the possibility of transfer. Distinguishing between superficial and structural features, the common belief is that transfer is dependent on the similarity between structural features. These define the causal relationship of an example and are central to its meaning (Holyoak, 1984). In contrast, superficial features are helpful in setting a problem’s context and can facilitate transfer, but their similarity between training and transfer is not a limiting factor.

Structural features determine if information can be applied between training and transfer tasks. Because structural elements define the meaning of a task, tasks that do not share structural features have different meanings, barring the transfer of information between them. There are two ways structural features can be defined. Gentner (1983; Gentner and Toupin, 1986; Gentner, Ratterman, and Forbus, 1993) defines elements as structural in terms of their relationships. An object is structural if it is an important component of the object-action relationships of a task. For example, in the analogy, “A hydrogen atom is like the solar system”, the sun and planets are structural elements of the solar system analog. Both are involved in rotating, attracting, and revolving around each
other; the most important procedures of the analog. In comparison, Holyoak (1984) defines structural elements by deciding if changes to certain features would also change the causal structure of the task. As will be shown later, this may be a more complete way of identifying structural elements.

An example will clarify Holyoak's definition of structural. Holyoak frequently uses Duncker's (1926) radiation problem as a training task. In the problem, a patient has a tumor that needs to be removed and surgery is not an option. A powerful X-ray can destroy the tumor, but it would also kill the surrounding healthy tissue. Doctors decide to use several weaker X-rays that simultaneously converge upon the tumor. In this way, the combined effects of the rays will destroy the tumor without harming the surrounding tissue.

The X-ray (and its assumed attributes) is a structural element. If the effect of the X-ray was not cumulative when projected from multiple directions, then the divide-and-converge strategy would not have been effective. Changing this attribute of the X-ray could alter the causal structure of the example, so the X-ray must be a structural element.

The fact that the X-ray could be split in to six or seven rays would be a superficial feature assuming that the difference in numbers did not affect the outcome of the story. Superficial features have two roles. Features that contribute to the context of a problem help to remind people of similar, previously-solved problems. An abundance of surface similarities leads people to repeat actions that were successful in the past. Reliance on superficial details to cue solutions is particularly evident with novices, as was shown by Chi, Feltovich, and Glaser (1981) in their comparison of physics experts and novices.
Understandably, part of the difficulty of surface features is that novices cannot easily
decide whether a feature is structural or surface in nature, or its role in the problem.

A second role of superficial features is to provide clues on how to map solutions
from one problem to another (Ross, 1987; Ross, 1989; Gentner, Ratterman, and Forbus,
1993). Gentner and Toupin (1986) presented children with stories in which the actors
were animals, one of which was a walrus. Children could more accurately repeat the
story if the walrus character was changed to a seal than if the character was a lion.
Adults show the same reliance on superficial details to help them apply old information
to a new problem (Ross, 1987, 1989). At a superficial level, it seems that object
similarity reduces cognitive load, simplifying the comparison of two examples.

Differentiating between structural and superficial features is an important step
towards understanding transfer. Identical elements do need to exist, but the elements
must be structural. Similarities between superficial features alone could create a situation
where a person recalls a superficially similar example that has little relevance to the
current problem. Even though two tasks may seem similar on the basis of superficial
similarities, being reminded of previous training examples is not useful unless the
training example also contains information that can be applied to the transfer problem.

Abstract Schemas

A theory exists that describes transfer in terms of structural elements. Some
researchers have applied the concept of schemas to the field of transfer (Reed, 1993; Gick
Royer, 1979). The basic assumption is that people possess abstract schemas that
represent a task’s structural elements with few of the task’s superficial elements. When a novel task is presented, a process of analogical mapping is performed that attempts to match the novel task with an abstract schema in memory. If a match is found, the schema is used to fill in gaps of information that exist.

The abstract schema theory predicts that transfer occurs when a given analog (i.e. example, task, or problem) has structural characteristics that match an abstract schema. Schemas that are abstract simplify the mapping process. Mapping between two non-abstract examples would require the comparison of structural similarities and the controlling of surface differences. Superficial characteristics act as noise, however, complicating the mapping process because they could be mistaken by the problem-solver for structural characteristics. Given that an abstract schema and analog share structural features, mapping from the schema only requires the comparison of similar elements. As a schema becomes increasingly abstracted, a process dependent on experience, it will represent fewer superficial characteristics and be easier to apply to a wider range of situations. Accordingly, a significant concern of the theory is the process by which people induce schemas from a set of related examples.

In a series of experiments that demonstrates the effect of abstract schemas on transfer, Gick and Holyoak (1983) documented some of the minimum requirements for subjects to create a schema. The concept subjects were to learn took the form of a story written by Duncker (1926). In it, a doctor has a patient with a malignant tumor. The tumor must be removed or the patient will die, but surgery is not possible. The doctor realizes that a ray of radiation could be used to destroy the tumor, but a dose strong enough to destroy the tumor would also destroy surrounding healthy tissue. The doctor
decides to project the radiation from several different directions simultaneously so it converges on the tumor and destroys it, while preserving the surrounding healthy tissue.

A variety of stories were written which could all be solved using the same divide-and-converge strategy from the Radiation story. Untrained Control groups used the divide-and-converge strategy to solve these types of problems only 10% of the time. After reading the Radiation story, 29% of subjects answered a transfer problem using the divide-and-converge strategy. The percentage of transfer increased to 32% when the story was combined with a sentence that summarized the strategy.

The level of transfer was about the same if a pair of examples from the same domain (military or fire-fighting) were used during training, but a dramatic improvement came from subjects in two different conditions. One group of subjects was told the strategy and read two stories set in different backgrounds (one set in a military background and the other in fire-fighting). The other was given a diagram comparing several small arrows converging on a point versus a single large arrow and two stories from the same background. These groups showed rates of transfer of 62% and 61%, respectively.

Gick and Holyoak pointed to the increase in transfer that came with multiple examples and interpreted this to mean that the induction of schemas was dependent on a person’s ability to draw on the similarities between examples. Although single examples do not allow for comparison, multiple examples do, and as each additional example is added, it becomes easier for the problem-solver to identify common structural elements. Common elements are even easier to identify if the examples come from different domains because there would likely be fewer coincidental similarities. Their conclusion
was that with enough diverse examples, a person would be able to create a schema devoid of coincidental, superficial characteristics. He would then be left with an abstract structural road map that could be used to solve any problem that had a similar structure.

Schema abstraction does have some faults primarily associated with the role of superficial characteristics in transfer. Characteristics of the context and of the objects present in transfer tasks have been shown to influence transfer of both novice (Ross, 1987, 1989; Gentner and Toupin, 1986) and experienced problem-solvers (Medin and Ross, 1989). Medin and Ross point out that the consideration of superficial characteristics may not be irrational. There may be more than a spurious correlation between sets of superficial and structural features (e.g., in math, the presence of a vehicle may imply a distance problem). There is also evidence that connecting a current problem with problems solved in the past is a critical step in transfer, a particularly central role of superficial features. Often the applicability of previously learned information goes unrealized because the information was associated with a context different from the one being faced.

Still, the fundamental argument of the abstract schema position on transfer is that transfer depends on structural similarities and that abstract schemas are the method by which people represent these structural similarities. Superficial similarities are useful for cueing the connection between related sets of problems, but reminding alone is of no use if no structural relationships exist. Some examples of findings on far transfer help to illustrate the difficulty subjects (and experimenters) have in separating structural and superficial elements, the contribution of both types of features, and the general difficulty
of producing transfer. These examples also point out pitfalls of transfer implied by the abstract schema perspective.

**Examples of far transfer**

The most-often cited case of unsuccessful transfer is presented by Reed, Ernst, and Banerji (1974). The study is notable because the solution for the training and transfer tasks was not only similar, it was identical. Reed et al. is actually a good example of the requirements for structural similarity. This study was arguably an example of a situation where the superficial features were similar enough to fool the researchers into predicting transfer, but the actual structural features were different. It shows that structural features are better defined by causal relationships (Holyoak, 1984) than by common operators (Gentner 1983; Gentner and Toupin, 1986). It also demonstrates how difficult it is to predict the elements that will transfer between two examples. Reed et al. had hoped that transfer could be measured in terms of the number of moves required to solve two puzzles, but it turned out that transfer was better measured in terms of time and illegal moves.

Half of the subjects in this study were presented with a training puzzle that consisted of a river, a two-person boat, three Missionaries, and three Cannibals. The goal was to move all six people from one side of the river to the other with the constraint that the Cannibals never outnumber the Missionaries on either side of the river. The transfer puzzle also contained a river and a two-person boat, but the people were replaced with three Jealous Husbands and their three wives. The constraint of the Jealous Husbands puzzle was that no wife could be with another man without her husband also present. For
the other half of the subjects, the Jealous Husbands puzzle (JH) was the training task and the Missionaries and Cannibals puzzle (MC) was the transfer task.

If subjects were allowed to solve the training task once before approaching the transfer task, there was no decrease in the number of steps required to complete the puzzle. No transfer occurred even though the transfer task followed immediately after the training task and existed in an almost identical environment. Another trial was also made after telling subjects that the training and transfer tasks were related. There was sign of unidirectional transfer such that subjects trained with the JH puzzle solved the MC puzzle in less time and with fewer illegal moves (e.g. Attempting to have more cannibals than missionaries on one side of the river.) In terms of the total number of steps required to solve the puzzle, however, there was no transfer. Reed et al. (1974) hoped that allowing subjects to solve the puzzles twice would improve the possibility of transfer. They again found unidirectional transfer, with subjects solving the transfer task with fewer illegal moves and in less time, but only if trained on the Jealous Husbands puzzle and tested with the Missionaries and Cannibals puzzle. The number of steps required to solve the transfer puzzle still did not decrease.

The improvement in terms of time and illegal moves does show that subjects transferred their knowledge of the JH task to the MC task. But, combined with the lack of transfer from the MC to JH puzzle, the findings imply that the MC and JH tasks may not be structural analogs. It could be that subjects transferred the information they believed applied to both puzzles, information that consisted only of rules regarding legal moves and heuristics that sped up their performance.
Reed et al. (1974) believed this study was a clear example of far transfer because only the surface features of the two puzzles were different; both puzzles could be solved using the same 11 steps. Even so, solving the training puzzle did little to decrease the number of steps required to solve the transfer puzzle. Beyond providing examples of successful and unsuccessful transfer, this study shows that structural similarity may be perceived by subjects in conceptual terms instead of in mechanical terms. Comparable to Holyoak’s (1984) definition of causal structures, the reason why missionaries and cannibals or wives and husbands needed to be separated or kept apart defined the structural relationship of the problems. The ability for missionaries and wives both to ride in a boat was not enough to make them structural analogs for each other. This study was an example of two problems with similar surface features and dissimilar structural features. Structurally, different causal reasons drove the solution of the MC and JH problems.

**An example of abstract representations in use**

The fundamental argument in favor of the abstract schema perspective is that abstract representations are easier to compare than concrete representations. Gott, Hall, Pokorny, Dibble, and Glaser (1993, in Detterman) reiterate the point. Gott et al. observed a group of expert technicians transfer electronic error-diagnosing skill to a completely novel electronics system. Before the transfer task, each technician was given the opportunity to work with the new system and determine the discrepancy between his area of expertise and the new system. The technicians who could represent information abstractly performed better than those who couldn’t.
The language used to frame questions illustrates the fundamental differences between high-transfer and low-transfer technicians. The better performers talked about routines and equipment in abstract terms. "A test station is a test station... The electronic test process is the same. Most differences are in the physical aspects of the equipment." (p. 267). In contrast, less proficient performers were less flexible. "I treated the Autos stations as new equipment -- unrelated to the Manuals stations." A higher level of abstract conceptualization existed for the more successful technicians across all the information types measured by Gott et al.: "How-it-works" knowledge, "How-to-decide-what-to-do-and-when" knowledge, and "How-to-do-it" knowledge.

Abstract representations allow problem-solvers to utilize pre-existing information. The perceived differences of the novel electronics system dissuaded less successful technicians from applying procedures and information they had learned in the past. In contrast, the more successful technicians were in a better position to transfer previous knowledge because they compared the electronics systems in abstract terms and found similarities.

The Law of Large Numbers and Far Transfer

The work of Richard Nisbett and his associates has produced the most consistent set of studies demonstrating far transfer (Fong and Nisbett, 1991; Fong, Krantz, and Nisbett, 1986). Students frequently transfer the concept of the law of large numbers (LLN) to multiple problem contexts different from the ones present during training (Fong et al., 1986.) In one experiment, students were taught the LLN using an abstract description, a concrete example, or a combination of the two. All three conditions
exhibited transfer, with the combined group outperforming the other two. Fong et al. also telephoned 193 students from an introductory statistics course under the pretext of an athletic poll. The half of the class polled after finishing the course was significantly more likely to base their answers on statistical reasoning than the half polled before taking course.

Fong and Nisbett (1991) followed this up by examining the effect of different training domains on transfer. Subjects were taught the LLN using three examples, all based on one of two different domains. A transfer test occurred either immediately after training or after a two-week delay and consisted of questions from both domains. If transfer were based on memory for specific details of the training examples, subjects would be expected to perform better on questions from the Same Domain.

Instead, subjects in both training groups were able to answer Same Domain and Different Domain questions equally well, and both better than the no-training, control group. The two-week delay was related with much lower scores on Different Domain questions, although it had little effect on Same Domain questions. (See Figure 1.)
Figure 1: Although performance deteriorated more for far transfer problems after two weeks, subjects who received training still scored higher than the no-training Control group.

The format of Nisbett’s studies provides a good benchmark for transfer research. His training methods incorporate the optimal blend of examples predicted by the abstract schema perspective, a combination of abstract principles and concrete examples from multiple contexts, and the LLN is a principle that allows for numerous, realistic variations. Unlike tasks such as the Towers of Hanoi, with unusual derivations such as the Monsters and Globes, realistic puzzles based on the Law of Large Numbers are relatively easy to create. The perceived validity of LLN questions makes them somehow more insidious examples of far transfer because their relationship with each other is less evident. Law of large numbers questions can be set in any number of contexts, all of which have optimal solutions based on the law of large numbers.

Why not teach abstract concepts?

There is a temptation to teach abstract concepts directly. A direct route to understanding abstract schemas would deliver the benefits of expertise without the time invested in solving examples. Unfortunately, there is reason to believe that abstract information is not useful in the absence of concrete examples. Several studies have looked at the effect of teaching abstract information directly (Gick and Holyoak, 1983; Cheng, Holyoak, Nisbett, and Oliver, 1986; Ross, 1987, 1989; Reeves and Weisberg, 1994), all reaching the same conclusion. Without concrete examples, people are unable
to apply the principles present in abstract examples. Although a principle might sound useful and relevant, without constraints, there are too many ways to implement an abstract rule for it to be useful. (See Figure 2 for rules that sound affective, but are too general to apply.) Concrete examples are needed to help people map an abstract representation to the current problem (Cheng et al., 1986). The superficial characteristics of concrete examples serve as rules for the application of a principle, providing constraints and implied directions for use. Greeno, Smith, and Moore (1993) emphasize the importance of experience grounded in a particular environment. They hypothesize that experience may take the form of understanding the affordances and invariants of a situation. Training includes more than a concept, students also learn the kinds of actions and repercussions that are possible given an situation. Although an end goal is for people to represent problems abstractly, they need to abstract these schemas on their own, based on their own sets of experiences.

1. **Total Picture:**

   Before you attempt a solution to a problem, avoid getting lost in detail…

2. **Withhold Your Judgement:**

   Do no commit yourself too early to a course of action…

3. **Models:**

   Verbalize, use language to simplify the statement of the problem, write it down…
4. *Change in Representation:*

Problem solving can also be viewed as a change in representation…

(Others not shown.)

Figure 2: Although it may sound useful, abstract information is too general to apply. There are too many possible instantiations of an abstract concept.


This brings us to the method through which examples are presented to the student. The common assumption is to treat students as a passive receptacle for information, presenting examples without waiting for student feedback. However, this same assumption has also been associated with historically poor levels of far transfer. As an alternative, it is possible to create a situation where the lesson to be learned adapts to student feedback. This type of interactive training has been supported on the grounds that it leads to a better quality of training. On this basis, it may also lead to higher rates of transfer.

**Interactive Training**

Interactive training is an attempt to improve upon the classic Expository style in which the student is treated as a passive receptacle for information. It is driven by the
belief that students learn best when they are active participants in the learning process.

The philosophy that underlies Interactive training is either labeled under the older name of Discovery Learning (Hermann, 1969 for a review) or its newer incarnation, Learner Control. Borsook (1991) provides a list of several properties that well-designed interactive materials should support, each of which has been operationalized in different ways. (Table 1)

- Immediacy of response
- Non-sequential access to information
- Adaptability
- Feedback
- Options
- Bi-directional communication
- Interruptability

Table 1. Advantages of interactive teaching materials (Borsook 1991, cited in Schwier, 1993).

There are several reasons to believe that interactive training should be more effective than Expository training and consequently lead to higher rates of transfer. Proponents of non-Expository training believe that active learners are best suited to reconcile new information with their existing knowledge structures because information in non-interactive lessons can not be as well-tailored to each student (Wilson, 1996). Thus, training that promotes participation from the student is hypothesized to be more effective than an Expository style because it doesn’t force the teacher's understanding on
the student (Tobin, 1993). Placing the locus of control with the student increases student motivation, which may play a critical role in the transfer of learning (Kinzie and Berdel, 1990; Cronin and Cronin, 1992). Kersh (1964) hypothesized that motivated students would continue to think about principles even after the learning phase was over, testing the application of hypotheses in domains besides the ones encountered during learning. In addition, the process of solving problems independently could become a habitualized response. Students could learn to approach problems on their own, without help from a teacher. It is also believed that deeper levels of processing result from the high level of student involvement required to solve each problem (Milheim and Martin, 1991) and that student-organized information is easier to recall than teacher-organized information (Newmann, 1956; Gray, 1987.)

There are two particular attributes of interactive learning that should directly facilitate transfer. Typically, students taught through non-interactive demonstrations have difficulty separating the signal of the phenomenon from the noise (Roth, McRobbie, Lucas, and Boutonne, 1997). In contrast, interactive styles require students to exercise their knowledge of the concept, forcing them to develop and refine their representations of the principle being taught. This is equivalent to forcing them to create representations and schemas instead of repeating skills by rote.

The second benefit of interactive learning is the ability to test hypotheses. This may be the most significant. Feedback that students receive during training help them to revise their mental models regarding the concept to-be-learned. Placing the student in an active role allows them to verify beliefs regarding the concept being taught and to receive evidence disconfirming their misconceptions. Both of these functions are known to be
important processes of hypothesis testing. The drawback of Expository training is that students often only receive information specifying confirmatory instances of a concept. The difference between receiving instances when a concept is true and combined feedback of true and false instances may seem subtle, but the addition of disconfirmatory evidence has been shown to improve the quality of hypotheses generated (Klayman and Ha, 1987; Ivancic and Hesketh, 1995).

The ability of students to receive disconfirmatory evidence addresses the instances when transfer does not occur. A lack of transfer does not necessarily mean that subjects did not learn the principle correctly the first time or that they could not recall it when needed (Keane, 1987), although these are the two functions that most interest transfer researchers. Instead, the subject might have recalled the solution but found it to be unsatisfactory or considered some other solution to be more applicable. Reimann and Schult (1996) call the process of selecting from appropriate hypotheses the Control Problem.

As part of problem-solving, subjects need to choose a problem operator that can apply to a given problem state. The hope of transfer researchers is that the operator chosen is the one previously taught in the training phase, and they tend to acknowledge only the use of that operator as a successful case of transfer. However, it is possible that the subject may have previously experienced an alternative solution operator that is equally or more applicable in the same state. When multiple applicable solutions exist for a given problem, solution selection is a competitive process. Not only does training need to show subjects the relationship between a solution and a given problem, but training also needs to reduce the applicability of competing solutions.
Viewed in the context of the Control Problem, the importance of the hypothesis testing aspect of interactive training is evident. It is only by experiencing the inappropriateness of some hypotheses that subjects can reduce the likelihood that those hypotheses will be used again in the same situation. In addition, reducing the applicability of one hypotheses increases the applicability of others, among which is hopefully the correct solution.

The Use of Interactive Simulations

Other studies have supported the use of simulations as interactive training tools. Of the two mentioned below, only one actually makes a comparison between interactive training and a non-interactive method. The other study is based on observations, which reaffirms the need for empirical data on the merits of interactive training.

A set of interactive tools was the focus of training for a study intended to familiarize students at a college for teachers (Woodrow, 1995.) The course was based around several simulators, an example of which was one that modeled objects in planetary orbit. Students were able to test the parameters necessary to place the Space Shuttle into a stable orbit around the earth. Although this study did not make comparisons between interactive and non-interactive mediums, the students who participated did provide several comments that resonate with the claims of interactive training.
"What intrigued me most was what happened when the resulting data didn’t agree with what I expected. This caused me to wonder why this discrepancy occurred. Curiosity and intrigue resulted from my exposure to this technology..." (pg. 158)

"I spent a lot of time on my projects mostly because I worked alone... I actually think I enjoyed working alone because then I got into lots of trouble and I had to find my way out myself. I learned a lot trying difference things by myself." (pg. 165)

Of course, the intention of this class was not to teach science, specifically, but to teach the use of technology in the classroom. Still, in rough sense, the course was successful in that it generated interest on the part of its students. The second quote is particularly encouraging because it is evidence of practice in hypothesis testing and problem-solving.

In a different study, Rivera-Perez (1996) trained subjects to monitor and maintain a healthy level of blood sugar using a program that simulated the glucose levels of a diabetic patient. The subjects were trained to adjust its blood sugar level by adjusting the simulated patient’s diet, level of exercise, and insulin. The novel approach Rivera-Perez took to create a comparison group was to record the actions of each subject in the Interactive group, and then play them back for a corresponding subject in the comparison group. Because each comparison subject watched the recording of one Interactive subject’s experience with the simulated diabetes patient, the two groups were presented
the same information and for roughly the same amount of time although one group had
the benefit of interaction and the other did not. Both immediately after training and after
a one week delay, the Interactive group scored higher on a series of diabetes tests
demonstrating the advantage of interaction.

Transfer and Interactive Training

This study is an attempt to replicate the far transfer study of Fong and Nisbett
(1991) with the addition of an interactive training condition. Although Fong and Nisbett
showed successful transfer, the performance of their subjects was qualitatively low. On a
scale from 1 (non-statistical) to 3 (statistical, and referred to the Law of Large Numbers),
students in the most successful condition across three experiments had a mean score of
1.9, with 2 signifying “a poor statistical response.” It is possible that interactive training
could lead to higher levels of learning and to higher transfer scores. A replication of their
findings from outside of their laboratory would also strengthen their claim that far
transfer can be produced reliably.

In the context of the abstract schema perspective of transfer, the main goal of
training is to promote the induction of an abstract schema that subjects can later use when
their knowledge of a novel domain is poor. To be successful, the training procedure
needs to help subjects differentiate between structural and surface features and impart an
understanding of the causal relationship of the structural features. The interactive
training procedure should also allow subjects the opportunity to test hypotheses regarding
the LLN.
The interactive training mechanism used in this study is a simulation set in a card-gaming context. The game is provided as the concrete illustration for the law of large numbers, following an abstract, written definition. In order to emphasize the structural features of the scenario, the game has a phase in which the subjects can manipulate the value of its structural features and observe changes in the game's outcome. The effect of the superficial features is implied to be minimal because the value of the outcome depends entirely on the interaction of structural features. It is hoped that the emphasis on structural features will help subjects to induce an abstract schema. The second phase of the simulation requires subjects to test the knowledge they have gained. Given the outcome and all but one of the structural values, the subject's task is to determine the missing value. Subtly shifting the perspective of the problem should force subjects to develop their understandings of the law of large numbers, instead of exercising them by rote.

There are two main hypotheses of this research:

1. Subjects who are trained in the law of large numbers will score higher on far-transfer tests than subjects who do not receive training.

2. Subjects in the Interactive-Simulation condition will score higher on the transfer tests than subjects in the Expository condition.

There are two other points to consider that are commonly associated with transfer. The first is the interaction between cognitive ability and training style. There are two sets of conflicting predictions. One is that students with high cognitive ability perform better
in less-structured learning environments (Zimmerman and Sassenrath, 1978) or that low cognitive ability students perform worse in less-structured environments (Veenman and Elshout, 1995.) Consequently, subjects with greater cognitive ability should show higher rates of transfer in combination with Interactive training than with Expository training. The alternative view on cognitive ability predicts that Interactive training provides additional information to low-cognitive ability students that they would not receive from an Expository style. This view predicts that low Cognitive Ability students will benefit more from Interactive training than from Expository training, while high-Cognitive Ability students will be unaffected by training type (see Atlas, 1996 for a summary). Both views are possible, so no hypothesis will be made regarding the effect of Cognitive Ability.

The second point concerns the resiliency of the training types. Previous studies have shown (see Bjork and Schmidt, 1992 for a review) that variability during training frequently leads to higher rates of long-term retention. It is expected in this study that the more variable Interactive-Simulation group will recall more of their training after a delay than the Expository group. Hence, the Interactive-Simulation group will have higher transfer scores after the delay than the Expository group.

An additional hypothesis is:

3. The Delay will have less of an effect on the Interactive-Simulation group than on the Expository group.
Method

Participants

There were 94 subjects who participated in groups of up to five. They were recruited from a pool of undergraduate students enrolled psychology classes at Rice University and given partial credit for their participation.

Procedure

Subjects were assigned to three Training Types: Interactive-Simulation (IS), Passive-Simulation (PS), and Expository (E). There was a fourth, no-training Control group. Together, the combination of the three Training Types and the Control group measured different aspects of the Training Type variable. The three Training Types were further divided in half into those who were tested immediately following training or after a one-week Delay. The Control group was always tested immediately, so was excluded from the Delay. Training Type and Delay were both between-subjects variables. The overall design was three by two with a floating control group. There were seven groups altogether.

At the start of the experiment, subjects’ cognitive abilities were measured using the Wonderlic Personnel test. Subsequently, the Training Types read an abstract definition of the law of large numbers (LLN) and a story describing scientists measuring a population of birds that demonstrated the law of large numbers (see Appendix A). The three training groups differed as to what they received for their third training example. The Interactive-Simulation group worked with a card-dealing simulator that was designed to teach the LLN, each subject in the Passive-Simulation viewed a recording of a
different member of the IS group, and lastly, the Expository group read a story set in
the same card-dealing context as the simulator (Appendix D.)

Following training, subjects in the Immediate condition were given the transfer
questions. Those in the Delay condition were thanked for their participation and asked to
return in a week's time.

Materials

*The Card Simulator:* The card simulator was an imaginary casino game involving
two types of card decks: a Cheater's deck and an Honest deck. Subjects in the
Interactive-Simulation group interacted directly with the simulator. Subjects in the
Passive-Simulation group watched a recording of the simulator, while no subjects from
the Expository group saw the simulator in action. The instructions to the simulator stated
that the Cheater's deck was missing the Ace (considered low) through 5 from two suits
while the Honest deck contained all 52 cards. The average value of the two decks was
explicitly stated. Counting Aces as a point and Kings as 13 points, the average value of a
Cheater's deck was eight and that of an Honest deck, seven.

Initially, Interactive-Simulation subjects were given complete control over the
variables of the card simulation for three minutes. They could manipulate major
variables: the type of deck, the number of hands they wished to see dealt, and the number
of cards to be dealt in each hand. Mean values for each hand were displayed on-screen so
subjects could observe the pattern produced by different combinations of their selections
(see Figure 3). Subjects were explicitly instructed to observe the different means caused
by larger and smaller samples and to notice the predictive benefit of increasing the
number of hands dealt as these were both manners through which the simulator
provided examples of the law of large numbers.

Figure 3. In phase 1 of the simulator, subjects can choose the Type of deck, Number of
hands dealt, and the number of Cards dealt per hand.

Following three minutes of complete control, the IS group was asked to use their
experience with the Cheater's and Honest decks to help them identify 20 unknown decks
(see Figure 4.) Each unknown deck was randomly decided to be Honest or Cheater's,
and the number of cards to be dealt in each hand was also randomly set to be either two
or six. The subject could deal hands one at a time for $5 each, with the average value of
each hand recorded on the screen. At any time, the subject could guess if the deck was
Cheater's or Honest. Each correct response earned a fictional $100 and each incorrect response cost $50. As the Interactive-Simulation group interacted with the simulator, the computer made a recording of their actions that could be played back later.

The card simulator was meant to illustrate the law of large numbers through the combined effect of the number of cards in a hand and the number of hands dealt. Six-card hands were more representative of the type of deck than 2-card hands, and a larger number of hands was more representative than a smaller number. However, the benefit of larger samples was most evident when a small number of 2-card hands was compared with a large number of 6-card hands. It was hoped that, if subjects had not already learned this from the previous 3-minute training, this lesson would be made clear to them as they tried to improve their ability to predict the type of deck from which the cards were being dealt. Through repeated interactions with the simulator, the subject could learn that dealing more hands increased his ability to choose the deck-type, especially if there were only two cards in each hand.
Figure 4. In phase 2 of the simulator, subjects could deal hands until they were ready to guess if the deck was Honest or Cheater.

Testing Materials. Eight questions were presented by computer (see Appendix B). They were all answerable by reasoning with the law of large numbers. Five of the questions came from Fong et al. (1987). Fong et al. had identified three question types, so questions were borrowed from these categories. One question was Probabilistic, two Objective, and two Subjective. A second Probabilistic question was borrowed from Tversky and Kahneman (1974; Appendix B.5). Each category of question differed in its outward representation of statistics. The concept of randomness was most clear in Probabilistic questions. The variation in sample outcomes was either stated or an
explicit, random generating device was provided. Objective questions required that
subjects make judgments on objectively sound data, without cues regarding the
randomness of the data. In the third category of questions, subjects made decisions based
on data that was Subjective in origin, data derived from a person’s opinions. Fong et al.
ranked the Probabilistic questions as the easiest to answer, followed by the Objective and
then the Subjective questions. Two additional questions were written that emphasized
the variability of small sample sizes (Appendix B.4 and Appendix B.8). In pilot studies,
it appeared that subjects could learn that large samples provided better estimates than
smaller samples without realizing that smaller samples would therefore provide more
extreme estimates. These two “Law of Small Numbers” questions were added to test this
belief.

All eight questions were written in contexts different from the ones in which the
training examples were written. The common similarity was that all eight questions
could best be answered using reasoning based on the law of large numbers. Thus,
although the eight questions were superficially dissimilar, they were similar in terms of
their law of large numbers structure.

The dependent variable used in this study was based on each subject’s answer to
the eight transfer questions. Every subject provided eight answers. The system used to
score the answers is described below.

Coding System

The 3-point coding system used by Fong et al (1987) and Fong and Nisbett
(1991) was used here. The following is Fong and Nisbett’s definition of the three ratings:
"1 = an entirely deterministic response. Responses in the category included those in which the subject made no use of statistical concepts such as sample size, randomness, or variability.

2 = a poor statistical response. Responses in this category included some mention of statistical concepts, but the explanation was incomplete or incorrect. These responses contain one or more of the following characteristics: (a) the subject used both deterministic and statistical reasoning, but was judged by the coder to have preferred the deterministic reasoning; (b) the subject used an incorrect statistical principle, such as Gambler's Fallacy; and (c) the subject mentioned some statistical concept, such as luck or chance, but was not clear about how it was relevant.

3 = a good statistical response. Responses in this category made correct use of a statistical concept. Some form of the law of large numbers was used, and the sampling elements were correctly identified. In general, the subject was judged to have clearly demonstrated how the law of large numbers could be applied to the problem."

Two coders scored ten percent of the transfer problems. Their scores were identical for 78% of the questions.

Results
Results from the Card Simulator

During the third segment of their training phase, the Interactive-Simulation group trained using the card simulator for 20 rounds. An increase in the proportion of correct responses between the first five and last five rounds was predicted as a sign of learning. (Rounds one and two were treated as ‘practice’ rounds so the first five were actually rounds 3-7.) It was also expected that subjects would notice the relative variability of 2-card hands compared to 6-card hands and learn to deal more 2-card hands. No significant differences were found for the proportion of correct responses (see Figure 5) or the number of hands dealt between early and late rounds (see Figure 6).

![Percentage of correct responses](image)

Figure 5: Subjects were not better able to identify Honest and Cheater's decks in their last five hands than in their first five.
Figure 6: The number of 2-card and 6-card hands subjects chose to see did not change significantly between the first five and last five hands.

Another mode for describing the combinations of hands dealt and accuracy of responses is in terms of dollar amounts. When dealt 6-card hands, the Interactive-Simulator group won an average of $28 after choosing to see an average of 4.7 hands. When dealt 2-card hands, the group won $20 per-round after seeing 4.6 hands. These sums can be contrasted to the $25 expected value for a pure guess ($100 for a correct response - $50 for an incorrect response, divided by two responses.)

The earnings of the subjects didn’t appear to differ much from the expected value of a guess, but this was difficult to assess without a basis for comparison. It could be that there are no successful strategies for the card simulator, especially for 2-card hands, and that pure guessing is the best route. Given that subjects had only 20 rounds to develop strategies for the card simulator, they were most likely to develop simple strategies so two simple strategies were tested.
Figure 7: The two simple strategies provide a perspective on the difficulty of the card simulator. Although simple strategies could earn money, they were more successful for 6-card hands than 2-card hands.

An Averages and Extremes strategy were tested. The Averages strategy was based on the average value of the cards dealt. One to seven hands were dealt before a response was made. If the average was seven or lower (the average value of an Honest Deck), this Averages strategy predicted that the deck was Honest; if the average was eight or higher (the average value of a Cheater's deck), the strategy predicted the deck was a Cheater. This strategy was tested for 4000 repetitions for each number of hands from one to seven, for 28,000 rounds altogether. The optimal result of this strategy varied for 6-card and 2-card hands. The strategy earned the most money after dealing
three, 6-card hands ($42). For 2-card hands, dealing a single hand before guessing earned the most ($32.) The simulated earnings for both strategies are shown in Figure 7.

Similar to the first strategy, the Extremes strategy dealt a maximum of one to seven hands for 4000 rounds at each number of hands. It was different from the Averages strategy because it did not always deal the maximum number of hands. If the average value of the cards was five or lower after any deal, the strategy predicted an Honest deck. Similarly, if the average value of the cards was ever 10 or higher, the strategy predicted a Cheater’s deck. It proved to be slightly more successful than the Averages strategy. Under the Extremes strategy, the optimal number of 6-card hands to deal ranged from three to five, all with a projected earning of $45. The optimal number of 2-cards was either four or five, with an expected earning of $35.5

According to these follow-up simulations, simple strategies could win more than pure-guessing for 2-card hands. The result of the strategy simulator also identified a design flaw in the card simulator. One purpose of the card simulator was to condition subjects to deal more 2-card hands than 6-card hands, but according to the Averages and Extremes strategies, it was only worthwhile to deal the same number or fewer 2-card hands than 6-card hands. This could explain why subjects dealt the same number of 2-card and 6-card hands.

This flaw makes it difficult to draw conclusions based on either of the Simulation-based groups because it is difficult to determined what they learned about the LLN from their interaction with the card-simulator. The card simulator did provide examples of the variability of 2-card hands; attentive subjects could have noticed the wide range of average values 2-card hands relative to 6-card hands.
While designing and testing the strategies, the variability of the card simulator was also made evident. On average, subjects experience roughly 10 rounds of 2-card hands and 10 rounds of 6-card hands. In comparison, it took 4000 rounds for the effect of the strategies to become stable. At 10 rounds, the relationship between any strategy and its payoff was difficult to determine. Ironically, with only 10 rounds, the true shape of the strategies was too variable to determine (the sampling size was too small.) Subjects might have learned more about the LLN if they could also have extended the number of rounds for which they interacted with the simulator.

Overall, the training phase of the Interactive-Simulation group may have been too complex. With only a total of 20 trials to develop a strategy, or even to learn that 2-card and 6-card hands required different strategies, it is unlikely that their understanding of the card simulator environment progressed very far. A simplification of the simulator would be a logical change in a subsequent study. This and other changes to the simulator are explored later.

Factors of the Law of Large Numbers

Previous studies on transfer had shown that even minor changes to learned materials affect the likelihood of transfer. Consequently, an attempt was made to determine the structure of the skill being transferred to the most accurate degree possible. Instead of treating the eight transfer questions as individual dimensions of the LLN, a factor analysis with varimax rotation was performed to isolate common elements.

Partial correlations between the eight questions were used in the factor analysis. There was some concern that the type of training might affect certain questions in a
similar manner. These training-induced correlations were undesirable, so a partial
correlation of was performed controlling for Training Groups. The correlations are
shown in Appendix C.

Contrary to Fong et al.'s prediction, correlations between the eight questions in
this study did not support their Probabilistic, Objective, and Subjective factors of transfer.
Objective questions correlated more with Subjective questions than with other Objective
questions, while Subjective questions correlated more with Objective questions than with
other Subjective questions. (See Table 2.) Thus subjects did not respond primarily to the
presence of probabilistic cues.

<table>
<thead>
<tr>
<th>Question Type</th>
<th>Probabilistic</th>
<th>Objective</th>
<th>Subjective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probabilistic</td>
<td>0.328</td>
<td>0.177</td>
<td>0.121</td>
</tr>
<tr>
<td>Objective</td>
<td>.</td>
<td>-0.055</td>
<td>0.432</td>
</tr>
<tr>
<td>Subjective</td>
<td>.</td>
<td>.</td>
<td>0.217</td>
</tr>
</tbody>
</table>

Table 2: Questions did not load cleanly onto the three factors proposed by Fong and
Nisbett (1986).

The data provided by the eight questions was instead reduced to three factors
which can be thought of as the Small Sample, Human Performance, and Large Sample
factors. The Darts and Stocks questions loaded predominantly on the Small Sample
factor, while the Admissions, Auditions, and Tryouts questions loaded on the Human
Performance factor, and the Babies, Cars, and Slots questions loaded on the Large Sample factor. The exact factor loadings can be found in Table 3.

<table>
<thead>
<tr>
<th>Question</th>
<th>Small Sample</th>
<th>Human Performance</th>
<th>Large Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Admissions</td>
<td>0.361</td>
<td>0.619*</td>
<td>0.084</td>
</tr>
<tr>
<td>Auditions</td>
<td>0.179</td>
<td>0.659*</td>
<td>-0.124</td>
</tr>
<tr>
<td>Babies</td>
<td>0.369</td>
<td>0.008</td>
<td>0.649*</td>
</tr>
<tr>
<td>Cars</td>
<td>-0.132</td>
<td>-0.131</td>
<td>0.632*</td>
</tr>
<tr>
<td>Darts</td>
<td>0.805*</td>
<td>0.015</td>
<td>-0.044</td>
</tr>
<tr>
<td>Slots</td>
<td>0.092</td>
<td>0.119</td>
<td>0.699*</td>
</tr>
<tr>
<td>Stocks</td>
<td>0.661*</td>
<td>0.174</td>
<td>0.222</td>
</tr>
<tr>
<td>Tryouts</td>
<td>-0.390</td>
<td>0.737*</td>
<td>0.082</td>
</tr>
</tbody>
</table>

Table 3: All eight questions loaded onto three factors. Numbers with a "*" represent the questions that loaded heavily on each factor. Each factor was constructed by averaging the questions that loaded on it. (The Small Sample factor was an average of the two questions that loaded most on it.)

To ensure that the factors accurately categorized the questions, the same partial correlation was repeated, this time including the three factors as variables. In support of the three factors, they correlated most with questions that loaded on them (see Table 4.)
<table>
<thead>
<tr>
<th>Type</th>
<th>Small Sample</th>
<th>Human Performance</th>
<th>Large Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Admissions</td>
<td>0.271</td>
<td>0.700 *</td>
<td>0.125</td>
</tr>
<tr>
<td>Auditions</td>
<td>0.179</td>
<td>0.663 *</td>
<td>0.001</td>
</tr>
<tr>
<td>Babies</td>
<td>0.297</td>
<td>0.078</td>
<td>0.738 *</td>
</tr>
<tr>
<td>Cars</td>
<td>0.082</td>
<td>-0.040</td>
<td>0.643 *</td>
</tr>
<tr>
<td>Darts</td>
<td>0.794 *</td>
<td>0.106</td>
<td>0.139</td>
</tr>
<tr>
<td>Slots</td>
<td>0.119</td>
<td>0.134</td>
<td>0.691 *</td>
</tr>
<tr>
<td>Stocks</td>
<td>0.840 *</td>
<td>0.174</td>
<td>0.260</td>
</tr>
<tr>
<td>Tryouts</td>
<td>-0.090</td>
<td>0.690 *</td>
<td>0.029</td>
</tr>
</tbody>
</table>

Small Sample 1.000 0.174 0.248
Human 0.174 1.000 0.079

Performance

Large Sample 0.248 0.079 1.000

Table 4: This partial correlation matrix shows that the pattern of correlations supports the use of three factors. Once again, the numbers with a "*" represent the questions that loaded most on each factor.

The questions comprising each factor can be logically related. Questions loading on the Large Sample factor (the Slots, Babies, and Cars questions) required subjects to base an answer on the more consistent of two samples, a typical response for LLN
questions. For the Small Sample factor (which consisted of the Stocks and Darts questions), subjects needed to understand the ability of small samples to vary greatly. Answers to these questions required the subject to explain why smaller samples might lead to less-common events or to choose the smaller of two samples, expressly to achieve an unlikely event.

Human Performance questions (Tryouts, Auditions, and Admissions) required subjects to predict the future performance of an individual given a previous history. These questions were characterized by the type of incorrect answer frequently provided. Although all of the questions could best be answered using the LLN, some subjects stated that human performance could not be described in statistical terms. Instead, they would provide a list of factors that could potentially be responsible for a given pattern of behavior.

Altogether, the three factors accounted for 56% of the variance. The questions comprising each factor correlated more with other questions within the factor than with questions outside of the factor. (Table 5)

<table>
<thead>
<tr>
<th>Question</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Small</td>
</tr>
<tr>
<td></td>
<td>Sample</td>
</tr>
<tr>
<td>Small Sample</td>
<td>0.363</td>
</tr>
<tr>
<td>Human Performance</td>
<td>0.119</td>
</tr>
</tbody>
</table>

(Table 5)
Table 5: Questions comprising the three factors identified by the factor analysis correlated more within each factor than with questions outside of the factors.

Identifying these three factors of the LLN helps to simplify the process of LLN transfer. By reducing the number of items being compared, the factor analysis has simplified the process of determining the match between the training definition of the LLN and its variable implementations. Analyses of the data was performed using the three LLN factors as dependent variables.

The pattern of the results

The results of this study are divided into four segments: an analysis of overall transfer, followed by separate analyses of the Human Performance, Small Samples, and Large Samples factors. The same set of analyses applied to all four sets of data and was used in the same sequence. The Overall Analysis and the Human Performance factor were more complicated and contain additional follow-up analyses. The four segments have these analyses in common:

- First, the influence of Cognitive Ability was examined to determine if aptitude influenced the effectiveness of different training styles. In addition to the main effect of Cognitive Ability, the Cognitive Ability by Training Type and Cognitive Ability by Delay interactions are tested, as well as the Cognitive Ability by Training Type by Delay, three-way interaction. Cognitive Ability scores were
used as covariates when their effect was consistent across groups (no significant interactions.)

- Next the effect of Training was investigated. The average of the Training Types was compared with the Control group, each Training Type was compared against the Control group independently, and the difference between each pair of Training Types was tested.

- Lastly, the main effect of Delay was tested as well as the Training Type by Delay interaction. This analysis was somewhat integrated with the above test of Training. The Control group was not included in these analyses because it was only tested in an Immediate condition.

**Overall Transfer**

![Overall Transfer Chart](chart.png)
Figure 8: The groups who received training scored significantly higher than the no-training, control group.

![Effect of High vs. Low Cognitive Ability across Training Groups.](image)

Figure 9: There was a significant main effect of Cognitive Ability. The subjects are divided along their median score of 32 in this graph to simplify presentation.

The test of overall transfer used the average of all eight transfer questions as the dependent variable (Figure 8.) The effect of Cognitive Ability (see Figure 9) was significant, \( F(1, 76) = 20.650, p < 0.001 \), but there was no Cognitive Ability by Training Type interaction, \( F(3, 76) = 0.940, p = 0.427 \), no Cognitive Ability by Delay interaction, \( F(1, 76) = 0.500, p = 0.483 \), and no Cognitive Ability by Delay by Training Type interaction, \( F(2, 76) = 1.520, p = 0.224 \). Because the effect of Cognitive Ability was consistent across groups, Cognitive Ability scores were used as a covariate in subsequent tests.
Subjects who received training scored higher than the no-training control group on the transfer questions. The average score of trained subjects (M = 2.106) was significantly higher than that of the untrained Control group (M = 1.778), F(1, 85) = 19.540, p < 0.001. Each training group was also compared to the Control group using a Dunnett’s test, and all three training groups’ scores were significantly higher than the Control group’s, replicating Fong and Nisbett’s findings that far transfer is possible, and showing that an interactive training technique can promote transfer. A Tukey test did not detect significant differences between the three training groups.

![Effect of Delay for Overall Transfer scores](image)

Figure 10: There was a slight decrease in scores after a one-week delay, but the difference was not significant.

Subjects tested Immediately after training scored slightly higher (M = 2.197) than those tested after the Delay (M = 2.008; see Figure 10), but this difference was not
significant, $F(1, 82) = 1.200$, $p = 0.277$. There was no Training Type by Delay interaction, $F(2, 84) = 0.870$, $p = 0.423$.

The Human Performance factor

![Transfer scores for the Human Performance factor.](chart)

Figure 11: Transfer was successful for the Human Performance factor. Subjects who received training scored higher than the Control group.

The main effect of Cognitive Ability was marginally significant, $F(1, 77) = 3.77$, $p < 0.056$. The Cognitive Ability by Training Type interaction was also nearly significant, $F(3, 77) = 2.410$, $p = 0.073$. Neither of the other Cognitive Ability-related interactions was significant: Cognitive Ability by Delay, $F(1, 77) = 0.750$, $p = 0.390$, and Cognitive Ability by Delay by Training Type, $F(2, 77) = 1.350$, $p = 0.265$. Because none
of the Cognitive Ability interactions was significant, Cognitive Ability scores were used as a covariate.

There was transfer of LLN knowledge for the Human Performance problems (Figure 11.) The Training Groups ($M = 2.000$) scored significantly higher than the Control group ($M = 1.605$), $F(1, 83) = 11.47$, $p < 0.001$, and each group also scored significantly greater than the Control group, according to a Dunnett’s test. A Tukey test did not detect differences between the Training Types.

Similar to the Overall Analysis, the Delay did not significantly affect the Training Types, $F(1, 83) = 0.470$, $p = 0.494$. There was also no Delay by Training Type interaction, $F(2, 83) = 1.670$, $p = 0.198$

A brief digression to examine the Cognitive Ability by Training Type interaction
Figure 12: The Cognitive Ability by Training group interaction almost significant for the Human Performance factor. (Cognitive Ability was treated as a continuous variable in the analysis, but as a dichotomous variable for to make it easier to graph.)

Because the Cognitive Ability by Training Type interaction was nearly significant, the analyses were repeated without using Cognitive Ability as a covariate. Instead, the analyses were repeated for High-Cognitive Ability and Low-Cognitive Ability groups separately. Subjects were assigned to either group based on a median-split of 32. There were 45 Low-Cognitive Ability subjects and 47 High-Cognitive Ability subjects.

The labels of High-Cognitive Ability and Low-Cognitive Ability are relative to this group of subjects. To put the scores in perspective, the Low-Cognitive Ability subjects had an average score of 26 while the High-Cognitive Ability scores had an average score of 35. The national average for the Wonderlic Personnel Test is 21, with a standard deviation of 4.

The nearly significant Cognitive Ability by Training Type interaction indicates that the slope of the relationship between Cognitive Ability and transfer differed as a function of the Training Type group. Although this finding was not statistically significant, it does provide an interesting trail to follow (see Figure 12.) It appears that the difference between the High-Cognitive Ability and Low-Cognitive Ability subjects was small for simulation-based training, but was large with no-training or with Expository training. The slope of the regression line was \(-0.009\) for both the Interactive-
Simulation group and Passive-Simulation group, but seven times higher for the Control group at 0.065, and six times higher for the Expository group at 0.056.

An alternative perspective on this interaction is that simulation-based training may have equalized the difference between High-Cognitive Ability and Low-Cognitive Ability subjects, leading to similar rates of transfer regardless of the aptitude of the subject.

High Cognitive Ability subjects showed no significant effect of Delay, $F(1, 41) = 0.030, p = 0.855$, and no Delay by Training Type interaction, $F(2, 41) = 0.050, p = 0.955$. However, according to the Dunnett’s test, the Expository group was the only group to score significantly higher than the Control. The average score of the three training groups was only marginally higher than the Control group, $F(1, 41) = 3.69, p = 0.062$, and the Tukey test did not detect any differences between the Training Types.

For the Low-Cognitive Ability subjects, the effect of Delay was not significant, $F(1, 37) = 0.260, p = 0.613$, and the Delay by Training Type interaction only approached significance, $F(2, 37) = 2.66, p = 0.084$. As Figure 12 shows, Interactive-Simulation and Passive-Simulation were very effective methods for training Low-Cognitive Ability Subjects. The average of all three training groups was significantly higher than the Control group, $F(1, 43) = 11.750, p = 0.001$, and the Dunnett’s test found significant differences between the two simulation-based groups and the Control group, but not the Expository group and the Control group. According to the Tukey test, there were no significant differences between the Training Types. It does appear that simulation-based training equalizes the difference between Low and High-Cognitive Ability students.
The Small Sample factor

The effect of Cognitive Ability was significant for the Small Sample factor, $F(1, 77) = 12.130, p < 0.001$, but there was no Cognitive Ability by Training Type interaction, $F(3, 82) = 0.93, p = 0.431$, no Cognitive Ability by Delay interaction, $F(1, 77) = 1.11, p = 0.296$, and no Cognitive Ability by Delay by Training Type interaction, $F(2, 77) = 1.760, p = 0.179$, so the Cognitive Ability was used as a covariate in this analysis.

![Transfer Scores for the Small Samples factor.](image)

Figure 13: The two simulation-based groups scored higher on the transfer test, but the difference was not significant.

There was no effect of Delay, $F(1, 83) < 0.00, p = 0.978$ and no Delay by Training Type interaction, $F(2, 83) = 0.240, p = 0.785$.

Although transfer was again successful, Small Sample questions were particularly difficult. Subjects who received training ($M = 1.769$) scored significantly higher than the Control subjects ($M = 1.463$), $F(1, 83) = 5.32, p = 0.024$, but no individual training group
scored higher than the Control group. The performance of the Control group shows how hard these questions were to answer statistically (see Figure 13). Untrained students provided answers with an average of 1.46, which is only slightly higher than a completely deterministic response of 1. Training was able to improve the conceptualization subjects had regarding highly variable small samples and the LLN, although no single style proved to be effective on its own. A Tukey test did not detect significant differences between the training groups. Comparisons did not show significant differences between the IS and PS groups, $F(1, 83) = 1.48, p = 0.227$, between the IS and Expository groups, $F(1, 83) = 0.02, p = 0.903$, or between the IS and PS groups, $F(1, 83) = 1.12, p = 0.292$.

The Large Sample factor

As with the other factors, there was an effect of Cognitive Ability, $F(1, 77) = 9.75, p = 0.003$, but no Cognitive Ability by Training Type interaction, $F(3, 77) = 1.820, p = 0.150$, no Cognitive Ability by Delay interaction, $F(1, 77) = 0.050, p = 0.815$, and no Cognitive Ability by Training Type by Delay interaction, $F(2, 77) = 1.530, p = 0.223$. Cognitive Ability scores were used as a covariate in the subsequent analyses.
Figure 14: All of the training groups scored higher than the Control group on the Large Sample factor.

The Delay did not significantly affect transfer scores, $F(1, 83) = 1.840, p = 0.179$; there was also no Delay by Training Type interaction, $F(2, 83) = 1.160, p = 0.317$.

Subjects who received training ($M = 2.421$) scored higher than subjects who didn’t ($M = 2.161$), $F(1, 83) = 5.230, p = 0.025$, but according to the Dunnett’s test, no single training group scored significantly higher than the Control group. The training groups also did not score significantly differently from each other.

Transfer scores were higher for the Large Sample factor than they were for either of the other two factors, but higher scores may have been a result of an inherent quality of this factor, rather than the influence of a training style. Without the benefit of training, subjects in the Control group still managed to use some statistical concept as part of their answer for these questions. The average score for untrained subjects was a fairly high
“2”, representing a poor statistical response (see Figure 14). It is likely that subjects were most familiar with the type of judgment required by Large Sample questions, since choosing an accurate sample is a skill that can be called upon across a wide range of situations.

A Comparison between the Three Factors

![Bar chart showing average transfer scores for each training type on each LLN factor.]

Figure 15: Large Sample questions were easiest for subjects to answer followed by the Human Performance and Small Sample questions.

Based on the performance of the Control groups, it appeared that Large Sample questions were the easiest type to answer, followed by Human Performance and Small Sample questions (see Figure 15). As a measure of untrained performance, the Control group was more successful in answering Large Sample questions than the other two
types, $t(26) = 4.200$, $p < 0.001$. This hierarchy between the factors was also reflected across the three training groups. Scores on the Large Sample factor ($M = 2.344$) were higher than those on the Human Performance factor ($M = 1.884$), $t(91) = 5.584$, $p < 0.001$, while scores on the Human Performance factor were higher than those on the Small Sample factor ($M = 1.679$), $t(90) = 2.237$, $p = 0.028$. The difference between the Large Sample and Small Sample questions was also significant, $t(90) = 7.898$, $p < 0.001$. These findings show that the three types of problems were not equally easy for subjects to answer. Instead, there is a hierarchy in which Large Sample questions are easiest to answer, followed by Small Sample and Human Performance questions.

**Discussion**

Overall, the successful documentation of transfer is valuable in that it supports Fong et al.'s (1986) findings and the more general claim that far transfer can be achieved. The argument for far transfer is particularly strengthened because the transfer questions were set in such a wide array of contexts, none of which repeated a context present during training.

Successful transfer to all three LLN factors is evidence in favor of the view that transfer occurs with little dependency on superficial similarities as long as common structural elements exist. As proponents of the abstract schema view would argue, the desired result of training in a skill or concept is a domain-independent representation that can be applied to multiple situations without being handicapped by superficial
differences. This study demonstrated that transfer does occur to multiple novel scenarios that share a structural element.

The advantage of Interactive training over Expository training was not supported, although there were signs of the advantage on the Small Sample factor. Although the effect wasn't significant, Interactive training led to higher transfer scores than Expository training on the Small Sample questions (Figure 13). The Passive-Simulation group also scored higher than the Expository group, suggesting that the effect was not caused by Interaction alone. A possible explanation for the higher success rate of the Interactive-Simulation and Passive-Simulation groups over the Expository group is that the lesson experienced by the simulation-based groups was more varied than the lesson experienced by the Expository group. While the Expository group read that the solution to the card simulator was based on the LLN, subjects in the simulation-based groups were given a chance to formed their own hypotheses regarding the effect of large and small sample sizes on the accuracy of a guess. In this way, subjects in the simulation-based training groups experienced a broader range of the LLN concept that could have helped them in later problems.

As was mentioned earlier, it is difficult to make observations based on the performance of the Interactive-Simulation group because of the complexity of the card-simulator. It is not certain what subjects were able to learn regarding the law of large numbers using such a complicated training tool. The original motive behind the design of the simulation was to develop a higher degree of understanding for the Interactive-Simulation group than the static, text description would for the Expository group. But, the complexity of the simulator may have confused the LLN principles it was intended to
demonstrate. At the least, the performance of the Interactive-Simulation group was on par with that of the Passive-Simulation and Expository groups.

There are some adjustments that could be made to simplify the use of the simulator. It is possible that subjects would have benefited from a detailed description of a distribution curve. The bulk of the information from the simulator was meaningful when interpreted in terms of the distribution curves of 2-card, Honest hands, or 6-card, Cheater’s hands (and the other combinations.) It may be that subjects did not have a structure for interpreting the data they were seeing and instead, relied on heuristics to determine if decks were Cheaters or Honest.

One other consideration is to alter the pay-off rules and eliminate the $5 fee to view each hand. This option was originally exercised to focus subjects’ attentions on the hands already dealt. There was a concern that without incentive to choose a deck type as soon as possible, subjects would spend too great a time dealing hands until they were absolutely sure of the outcome. Unfortunately, the $5 price caused its own complication because it transformed the simulation from a scenario in which a successful strategy dictated that “larger samples were more accurate”, into one in which a successful strategy needed to optimize the balance between the number of hands dealt and the expected values of the payoff. In this second case, the information gained from 2-card hands was less than that gained from 6-card hands, so the optimizing rule dictated that subjects were better off dealing fewer 2-card hands than 6-card hands. These two future changes may improve ability of the card-simulator to demonstrate the law of large numbers.

In terms of Cognitive Ability, there was a positive relationship with performance as predicted. No significant interactions between Cognitive Ability and Training Type
were found, although this interaction approached significance for the Human Performance factor. A graph of the scores (Figure #J3) showed that low-Cognitive Ability students in the Interactive-Simulation and Passive Simulation groups scored higher than low-Cognitive Ability students in the Expository or control groups. The implication is that simulation-based training provides the best form of training for low-Cognitive Ability students. The high-Cognitive Ability students scored highest after Expository training, supporting the point of view that the Expository style is best for smarter students. Still, this effect was not significant and it is difficult to determine the effect of Cognitive Ability in this study because of the restricted range of Cognitive Abilities represented. Overall, subjects scored three standard deviations above the national mean on Cognitive Ability, with a median score higher than the mean. Even the group considered “low-Cognitive Ability” had a mean score almost two standard deviations above the national mean. Realistically, this study did not represent a wide enough range of cognitive aptitudes for an effective comparison between low and high-cognitive ability groups.

The last point to consider is the effect of the delay. There was not a significant effect of Delay, either as a main effect or as an interaction with Training Types. The original reason for believing that the Interactive-Simulation group would perform better than the Expository group after a delay was based on Schmidt and Bjork’s (1992) findings that variability during training sometimes leads to greater retention. This study’s findings on the delay suggest is that the length of the Delay was not long enough to decrement memory of the training phase because neither group’s performance was significantly altered.
Three Factors of the Law of Large Numbers.

The identification of three factors suggests differences in subjects' conceptualizations of the law of large numbers. Two of the factors can be described as reflections of each other, namely the Law of Large Numbers (LLN) factor and the Law of Small Numbers (LSN) factor. The questions that were combined to create the LLN factor could be answered by choosing the sample that was larger. This is a common instantiation of the law of large numbers in the literature (Fong et al., 1986; Fong and Nisbett 1991; Tversky and Khanneman, 1974) and is probably one that is most useful in day-to-day experience. In contrast, the Law of Small Numbers questions required subjects to understand that small samples vary to greater extents than large samples. The third factor, Human Performance, was distinguished because some subjects responded that events dealing with human performance could not be explained using statistics.

The distinctions between the LLN factors are important because they illustrate general elements that can interfere with transfer. The first general observation is that students have a narrow interpretation of the rules they learn. The difference between the LLN and LSN questions is a good example. Given that large samples are more accurate, it is not a difficult logical step to derive that small samples are more variable. But, not all subjects performed this derivation. Instead, subjects were more likely to use statistical answers for LLN questions than for LSN questions.

The second observation is that recalling the correct answer is not the sole requirement for transfer. This was evident for the Human Performance problems. Theories of transfer commonly predict that transfer is dependent on a subject's insight
that a previous solution can apply to the current problem, but some subjects bypassed statistical data and constructed answers based on heuristics and information inferred from details in the scenario. A common example of a solution based on secondary information was to claim that the historically inferior quarterback in the Tryouts question (see Appendix B.1) might have enjoyed a remarkable summer-time improvement in skill. Although this explanation is possible and does fit the scenario, it is non-statistical because the likelihood that the inferior quarterback would permanently deviate from his historical performance is smaller than the likelihood that the superior quarterback is temporarily deviating from his historical performance.

Here are three responses that show how subjects undervalued the salience of statistical principles in comparison to their personal experiences:

1. To support why the historically proven actress in the Auditions question should not be chosen, one subject noted that, "The consistency in something like bird eyes [referring to the second training example] is not comparable to consistency in regards to such a creative endeavor."

In response to the question, "Did any of the examples help you answer the [transfer] questions?":

2. "Some... especially the ones with large survey sample. Often we must rely on intuition and common sense, though. Many of the above mentioned situations are hypothetical and depend on a person's mood, bias, etc..."
3. “Yes, when it came to those questions about odds that cannot be controlled like with the dice or with the slots. But those things that are human nature, like playing ball well or acting, circumstances can really have a huge effect on things. I don’t think you can really compare that to cards. People can have a bad day, but cards cannot.”

Solution Competition

Transfer may be just as dependent on the acquisition of new solutions as it is on the competition between new and old solutions. Transfer may not hinge entirely in terms of similarity-between-problem-states as analogical mapping tends to assume. Based on the results of this study, it is likely that transfer is also dependent on a competitive solution-selecting process, something like the Control problem brought up by Reimann and Schult (1996). The subject is presented with a problem space from which he identifies relevant elements. From the subject’s perspective, the subject’s goal is to choose a solution that sufficiently explains the effects of all relevant elements. It may be that similarity between surface features causes learned solutions to come to mind, but previous theories have not considered that multiple solutions may come to mind.

Human Performance problems are an excellent example of the competition that occurs between solutions. All of the Human Performance problems were designed so the optimal solution was based on the law of large numbers. In this sense, if subjects were interested in providing an incontestably correct answer, it would have to have been based on the LLN. Human Performance problems often led to non-statistical answers because the subjects themselves had experience performing exceptionally well or poorly. These
experiences intruded and were the seeds for solutions that incorporated anecdotally vague concepts like a "bad day" or "the right actress for the part". Personally experienced solutions were more salient than statistical ones, despite being less statistically sound.

If we assume transfer is a competition between solutions, some problem states can have a very good fit between the ideal solution and the problem. For the most part, Large Sample questions fall into this category. Given two different sample sizes and a requirement to choose the more representative sample, the correct choice is almost always to choose the larger sample. There are only a few characteristics that could be inserted that would make a LLN solution either incorrect or not salient; a larger but biased sample would make the LLN solution incorrect, and Human Performance considerations make LLN solutions not appropriate. Otherwise, it is unlikely that subjects would have preconceived rules that apply to this situation.

Other problem states, however, are not so clearly suited for a single type of answer. Duncker's x-ray scenario is one such problem. It is easy to assume that Gick and Holyoak (1983) were correct in stating that multiple examples were better than single examples because their experiment supported this claim. But, what if their single example had characteristics that made it a bad source for transfer? Even a minor change in one of several variables shifts the optimal solution away from the original multiple-directional-ray concept in the X-ray example. What if the subject considered chemotherapy as a cure for the tumor? Or orthoscopic, fiber-optic, laser surgery? In light of competing solutions, the one proposed by Duncker is only one among many, instead of being a
dominant response. Maybe subjects reading this story understood the solution as a leap of faith. It isn’t the best solution, it’s only the one offered by the ‘people in charge.’

Transfer between the X-ray scenario and the Generals scenario used by Gick and Holyoak (1983) faces the same problems, raised to another degree. If the original solution was not dominant, is there any reason why it should out-compete other possible solutions for the transfer problem? In the General scenario, a general needed to lead his army to capture a well-defended fortress. Unable to march his entire army down any single road, the solution was to divide his army and approach from several directions. Although this divide-and-conquer strategy is a possibility, once again, it is not demonstrably better than other solutions. What kind of mine destroys large numbers of people but not small groups? Wouldn’t dividing into smaller groups only increase the probability that more of the hidden mines would be found? Why not just remove the mines from one road and continue?

Transfer is a possibility between these two scenarios because they can share a similar solution, but the optimal solution to either may not have been the one they had in common. There is no way to know if subjects repeated the solution only because they read it in several examples and believed in it, or if its application was a demand characteristic.

Instead, transfer should be most likely when the optimal solution between two problems is the same. This is often confused with two problems having solutions in common, which are possible, but not optimal.
How does thinking of transfer as a Competitive process change the requirements for transfer?

In relation to the number of actions a person might associate with a given scenario, a competitive view acknowledges that subtractive forms of training are sometimes necessary. Typical tests of transfer have only added positive examples to subjects’ experiences, but not removed inaccurate preconceptions.

If a subject continuously experienced examples from Gick and Holyoak’s world of divide-and-converge, they would develop strategies in which dividing and conquering became a viable option. However, what if a group of subjects previously learned a strategy for dealing with despots fortified behind castle walls that relied on siege tactics? The original scenario provided by Gick and Holyoak did not specify that the invading army was under a time constraint to capture the despot, so either the divide-and-conquer or siege strategies could be applicable. Even though waiting for the despot to starve to death is a relevant response, Gick and Holyoak would have scored it as incorrect.

Subtractive training would be necessary to reduce the perceived effectiveness of some solution strategies, thereby increasing the relative effectiveness of the strategy-to-be-trained. In addition to learning why some strategies work, students need to know why certain strategies don’t work. If, for Human Performance problems, the training materials somehow illustrated how human performance could be estimated by statistical information, subjects may not have perceived responses based on anecdotal experience as being appropriate.
Overall, this study was successful in demonstrating the existence of far transfer for the law of large numbers. The results support the claim made by abstract schema theorists that transfer is fundamentally based on similarities between structural features. Although the method of interactive training used in this study was not shown to be significantly better than a passive style of training, improvements were suggested that could increase its effectiveness. The final issue discussed was conjectural, but worth considering. It is possible that effective transfer requires a teaching mind-set different from the 'additive' view historically supported. If solution selection is a competitive process, then students' ability to transfer skills might benefit as much from learning which actions not to take as from learning correct actions.
References


Appendix A: Law of large numbers examples.
Appendix A.1: Abstract example of the law of large numbers.

Investigators of human reasoning have found that some principles can improve judgement in situations of limited information. One of these is the Law of Large Numbers (LLN).

According to the LLN, a large sample randomly drawn from a population will be more representative of that population than a small sample. When applied in reverse, it suggests that a smaller sample will deviate more from the population than a larger sample.
Appendix A.2: Concrete example of the law of large numbers.

The LLN is often used to make educated guesses about some characteristic of a population when the entire population is not available, or when practical reasons prevent observation of the entire population.

Consider the researchers who track the grackels (those black birds) around Rice. (Many of grackels wear metallic, identification-bracelets.) Assume that the researchers are interested in all of the grackels who live within Houston city-limits, a number that they project to be around 3,000. These 3000 birds are the population they are interested in. The researchers want to know the percentage of grackels who have green eyes.

In actuality, 80% of the birds have green eyes, but this fact isn't available to the researchers. They need to devise a way of discovering it.

The researchers begin by discussing the possibility of catching every grackle within city limits and checking its eye-color. They quickly decide that they need a more practical procedure. Instead, they decide to catch a few birds and see what eye-colors they have.

They begin with a single bird and note that it has yellow eyes. So far, it appears that 100% of the birds have yellow eyes. Being experienced in statistics, they realize that one bird is a very small sample. According to the Law of Large Numbers, smaller samples are more likely to be different from the population than larger ones.

With a single bird, they would be underestimating the prevalence of green eyes by 80%.
They manage to catch four more birds. One bird has yellow eyes and three have green eyes. The researchers update their percentages to 40% yellow and 60% green-eyes. Still, the researchers realize that there are many hundreds if not thousands of grackels in the Houston area. Five is still a small sample, so not too dependable. At this point, their new percentages underestimate the prevalence of green eyes by 20%. Closer than before, but still off by a lot.

After a month of furious bird-catching, the researchers manage to document the eye-color of 195 more birds, for a total of 200 birds (including the five previous birds.) All together, they find that 82% of the grackels have green eyes. At this point, the researchers decide to stop. The sample is large enough that even if they caught 10 more birds, all with yellow eyes, the percentage of green-eyed grackels would only drop to 78%.

Citing the Law of Large Numbers, the researchers state that their sample is large enough that they are confident it accurately reflects the percentage of green-eyed grackels in the population. Unknown to them, 80% of the 3,000 grackels in the area they've chosen to research have green eyes, so the researcher's estimate of 82% is very good.

This example has shown that as the size of the sample increases, so does its ability to represent the population. With a very small sample, the researchers would have stated that all grackels in the Houston area had yellow eyes. After collecting larger samples, they were able to improve their estimate of the population.

Although it is always possible that a sample does not accurately represent a population, the possibility is lower for larger samples.
There is one last caution, however. Even though the Law of Large Numbers states that larger samples will be more representative of the population than smaller samples, this assumes that the samples themselves are randomly selected. As long as the researchers catch grackels from every part of Houston, or at least allow every grackle the same opportunity to be caught, their samples will follow the Law of Large Numbers.

Suppose the researchers did not know that eating high quantities of greasy foods led to yellow eyes in grackels, and they chose to catch birds around the dumpster-end of eating areas. In this case, their samples are biased in favor of birds with yellow eyes, so the Law of Large Numbers will be of less use. The researchers will have a tendency to catch birds with yellow eyes.

In the long-run, however, the Law of Large Numbers will still be valid. Once again assuming that there are 3000 birds in the area and that 80% of them have green eyes and 20% have yellow eyes, there are only 600 birds who have yellow eyes. As the researchers catch more and more birds, they will run out of birds with yellow eyes and begin to catch an increasingly large number of green-eyed birds. Even with a biased sample, the Law of Large Numbers will predict that larger samples will be more representative than smaller samples. Using biased samples, it will take larger samples.
Appendix B: Eight Transfer Questions

Appendix B.1: Tryouts

Coach Graves has been scouting two different quarterbacks for his college team, Mike and Scott. According to high-school records, Mike has performed excellently and seems to have a promising future. Scott's records show he has been a little less productive. Although he does show some promise, his history suggests that he will become a back-up quarterback.

During the two-day try-outs, Mike has difficulty completing passes, while Scott sails through easily. His assistants tell Coach Graves that Mike is performing poorly, while Scott is doing very well.

Coach Graves needs to choose the best quarterback he can get and he can only choose one. Which quarterback should Coach Graves choose and why?
Appendix B.2: Slots

Keith was driving through Nevada when he stopped in a gas station with two slot-machines. An old man nearby said, "There ain't no winning system for slot machines. It's all luck. You just put in a coin, pull the lever and hope you win. But let me tell you this: some machines are easier to lose on than others. That's because the owners can change the mechanism of the slots so that some of them will be more likely to make you lose. See those two machines? The one on the left will give you even chances of winning, but the one on the right is fixed to make you lose much more often than you win. Take it from me, I've played them for years."

Keith played both machines for a few minutes. On the left machine, he lost twice as much as he won. On the right machine, he won twice as much as he lost. He concluded that the old man was wrong about the odds of winning on the two slot machines; the opposite was true, the one on the right was more favorable than the one on the left.

Assume the old man has nothing to gain from the slot machines. Comment on Keith's conclusion and reasoning. Do you agree? Explain your answer.
The Caldwells were looking for a "safety-conscious, Swedish" car. As luck would have it, their old car stopped working on the last day of the closeout sale for the model year for both the Volvo and Saab. They quickly got out their Consumer Reports where they found that the consensus was that both cars were very sound mechanically, although the Volvo was felt to be slightly superior on some dimensions. They also found that readers of Consumer Reports who owned Volvo's reported having somewhat fewer mechanical problems than owners of Saabs. They were about to make a deal when Mr. Caldwell remembered that they had two friends who owned a Saab and one who owned a Volvo. Mr. Caldwell called up the friends. Both Saab owners reported having a few mechanical problems, but nothing major. The Volvo owner exploded when asked how he liked his car. "First that fancy fuel injector computer thing went out: $250 bucks. Next I started having trouble with the rear end. Had to replace it. Then the transmission and the clutch. I finally sold it after 3 years for junk."

Given that the Caldwells are going to buy either a Volvo or Saab today, which do you think they should buy?
At a bar, Mark and Joe are about to play one tie-breaker game of darts. Both know that Mark is better with darts, so they agree that Joe can choose the rules for the last game. Joe has to play by the same rules he chooses for Mark.

How can Joe maximize his odds of winning? Explain your reasoning.
Appendix B.5: Babies

There are two hospitals in Normaltown, TX; Central and Valley. At Central Hospital, 100 babies are born a day. At Valley Hospital, 30 babies are born a day.

Are days with more than 60% female births:

1) More common at Central hospital;
2) More common at Valley hospital;
3) Equally common at either hospital.

Explain your position.
Appendix B.6: Auditions

The director of a Broadway play just finished auditions for the female lead. Two of the candidates gave readings for the part that were very good. The third was given by an actress he had worked with in four previous plays. The director thought she had been superb in each. Unfortunately, of her four readings, one was very good, but the other three had been flat.

The director needs to choose the female lead immediately.

What should the director do? Hire the third actress or hire one of the two whose readings he like better? Why?
Appendix B.7: Admissions

The psychology department keeps records on the performance of thousands of its graduate students and relates this performance to scores on all kinds of background information about the students. Recently, there was a student from a small college with GRE scores and a GPA such that almost all accepted students had scores as high or higher, while rejected students typically had lower scores. The letters of recommendation were quite good, but none of the writers of the letters were personally known to the reviewers.

One of the reviewers argued against admission because students from small colleges tend to perform below the department average. Another disagreed, noting that one student admitted years ago from a small college also had similar scores, but became one of the top three students in the department.

Comment on the arguments put forth by the two reviewers. What are their strengths and weaknesses? Should the student be admitted and why?
Appendix B.8: Stocks

In the last five years, the stock market of a certain country has been in decline, with investors losing an average of 5% each year. Assuming that Person S had chosen to invest his money in a single company and Person D had chosen to invest his money in ten different companies, compare the performance of the two investors for the last five years.

How did the two investors perform?
Appendix C:

Partial correlation of transfer questions, controlling for training type.

<table>
<thead>
<tr>
<th></th>
<th>Admission</th>
<th>Auditio</th>
<th>Babie</th>
<th>Cars</th>
<th>Darts</th>
<th>Slots</th>
<th>Stock</th>
<th>Tryout</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>1.000</td>
<td>0.202</td>
<td>0.165</td>
<td>-</td>
<td>0.191</td>
<td>0.119</td>
<td>0.249</td>
<td>0.192</td>
</tr>
<tr>
<td>ns</td>
<td>0.061</td>
<td></td>
<td></td>
<td></td>
<td>0.045</td>
<td>0.056</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auditions</td>
<td>0.202</td>
<td>1.000</td>
<td>-</td>
<td>-</td>
<td>0.137</td>
<td>0.089</td>
<td>0.140</td>
<td>0.187</td>
</tr>
<tr>
<td>Babies</td>
<td>0.165</td>
<td>-0.045</td>
<td>1.000</td>
<td>0.105</td>
<td>0.180</td>
<td>0.323</td>
<td>0.281</td>
<td>-0.027</td>
</tr>
<tr>
<td>Cars</td>
<td>-0.061</td>
<td>-0.056</td>
<td>0.105</td>
<td>1.000</td>
<td>-</td>
<td>0.138</td>
<td>0.130</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.016</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Darts</td>
<td>0.191</td>
<td>0.137</td>
<td>0.180</td>
<td>-</td>
<td>1.000</td>
<td>0.095</td>
<td>0.324</td>
<td>-0.144</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.016</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slots</td>
<td>0.119</td>
<td>0.089</td>
<td>0.323</td>
<td>0.138</td>
<td>0.095</td>
<td>1.000</td>
<td>0.073</td>
<td>-0.014</td>
</tr>
<tr>
<td>Stocks</td>
<td>0.249</td>
<td>0.140</td>
<td>0.281</td>
<td>0.130</td>
<td>0.324</td>
<td>0.073</td>
<td>1.000</td>
<td>-0.020</td>
</tr>
<tr>
<td>Tryouts</td>
<td>0.192</td>
<td>0.187</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.000</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.002</td>
</tr>
</tbody>
</table>

Table C.1: These are the partial correlations of the eight transfer questions, controlling for Interactive group.
Appendix D:
Example three for the Expository group.

In the early days of gambling, a small Las Vegas casino had a game involving two types of card decks, Honest decks and Cheater's decks. Honest decks had 52 cards and an average value of 7. The Cheater's decks were missing the Ace through 5 from two suits. Because they were missing some low cards, Cheater's decks had an average value of 8.

The rules of the game were simple. The player tried to guess whether the deck currently in use was Honest or Cheater's.

Each game, the dealer would choose a new deck of cards and flip a coin. If the coin came up heads, the dealer would deal 2 cards. If the coin came up tails, the dealer would deal 6 cards. For $5, the player could also ask the dealer to deal another hand. The player could ask for as many extra hands has he wanted, each for $5. At any time, the player could guess whether the deck was Cheater's or Honest. If the player was right, he would win $100; if he was incorrect, he would lose $50.

When the player asked for another hand, the dealer would collect the cards, shuffle them, and deal as many cards as he had dealt the first time. If the original coin toss had come up heads and the dealer had dealt a 6-card hand, then every deal would be a 6-card hand until the player guessed whether the deck was Honest or Cheater's.
After years of playing this game, the casino finally stopped after one careful gambler figured out a winning strategy for the game. The gambler applied the law of large numbers to this game and won a fortune.

It turns out that the deck used each round could be thought of as a population; it was the group that the gambler wanted to know more about. The hands dealt by the dealer were samples from the population. Understanding that large samples were more like the population than smaller samples, the gambler knew that 6-card hands would be more like the deck they came from than 2-card hands. Each 6-card hand was much more likely to have an average value closer to the average value of the deck from which it was dealt than 2-card hands.

The game had another dimension, also. Because the gambler could ask the dealer to deal more than one hand, he could increase the number of samples he saw from the deck, thereby increasing his ability to predict the type of deck.

Using his knowledge of the law of large numbers, the gambler wrote this strategy for his friends:

1. 6-card hands are more representative of the deck than 2-card hands. If the deck is a Cheater's deck, 6-card hands will have an average value closer to the Cheater Deck's average value of 8. For Honest decks, 6-card hands will have an average value closer to the Honest Deck's average value of 7.
2. The more hands the dealer deals, the easier it will be to guess the type of deck. The total number of hands dealt can be considered a sample; the average value of each hand is one number in that sample. By dealing more hands, you will be able to increase the size of your sample and the accuracy of your guess. Bigger samples lead to better estimates of the population; of the type of deck being used.

3. Combining the first two rules, you should deal more hands when there are only 2 cards in each hand than when there are 6 cards. If it didn't cost $5 to deal another hand, you would deal many more hands, but since each hand costs money, you want to deal as few as possible. The bigger, 6-card hands have averages closer to the average of the deck, so a few hands will tell you a lot about what kind of deck the dealer's currently using. The average values of the smaller, 2-card hands aren't as consistent. Even if the dealer's using an Honest deck with an average value of 7, with a 2-card hand, you might get an average value of 5 one hand, and an average value of 12 the next hand.