

Abstract Knowledge and Reliance on Similarity in
Statistical Problem Solving

by

Angela Micco

B.A. University of British Columbia, 1986

M.A. University of Victoria, 1990

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Psychology

We accept this thesis as conforming
to the required standard

ACCEPTED

FACULTY OF GRADUATE STUDIES

DEAN

DATE

18 JUN 93

Dr. Michael E. J. Masson, Supervisor (Department of Psychology)

Dr. David F. Hultsch, Departmental Member (Department of Psychology)

Dr. D. Stephen Lindsay, Departmental Member (Department of Psychology)

Dr. C. Brian Harvey, ~~Outside Member~~ (Psychological Foundations in
Education)

Dr. Darrin R. Lehman, External Examiner (Department of Psychology)

© Angela Micco, 1993

University of Victoria

All rights reserved. Dissertation may not be reproduced in whole or in
part, by photocopying or other means, without permission of the author.

Supervisor: Dr. Michael E.J. Masson

ABSTRACT

Reminders -- the retrieval and use of examples from episodic memory -- have been characterized as a problem solving strategy indicative of individuals who do not understand the principle underlying a problem's solution (Ross, 1984). Whereas past research has provided insight into how learners in a new domain notice and use examples, the question of whether the use of examples continues after the individual has acquired an abstract understanding of the problem's underlying structure has not been adequately addressed.

In Experiment 1, subjects were differentially trained such that half developed an abstract understanding of elementary probability principles, and half did not. Moreover, the existence of the knowledge difference was demonstrated. Similarly in Experiment 2, subjects learned pragmatic inferential reasoning rules, and evidence of rule acquisition was demonstrated. In both experiments, evidence that individuals who understood the principle underlying the problem's solution nonetheless solved the problem by analogy to an earlier example was demonstrated by the emergence of a negative transfer effect. That is, subjects who understood the problem's underlying principle were more likely to use an inappropriate solution procedure when the test problem's story line reminded them of a training problem that used a related but different principle, than when the test problem's story line was new to the experiment. Furthermore, the results of Experiment 1 indicated that memory of an earlier example also influenced how individuals who

understood the problem's underlying structure applied the principle to the test problem. The results are discussed in terms of the use of a heuristic by which problems appear similar on the surface are solved using the same solution procedure.

Examiners:

Dr. Michael E. J. Masson, Supervisor (Department of Psychology)

Dr. David F. Hultsch, Departmental Member (Department of Psychology)

Dr. D. Stephen Lindsay, Departmental Member (Department of Psychology)

Dr. C. Brian Harvey, Outside Member (Psychological Foundations in Education)

Dr. Darrin R. Lehman, External Examiner (Department of Psychology)

Table of Contents

Abstract	ii
Table of Contents	iv
List of Tables	vi
Acknowledgement	vii
Dedication	viii
Chapter 1	
Introduction	1
Reminders in Problem Solving	4
Noticing Examples of Past Problems	5
The Use of Reminders	
in Problem Solving	7
The Differential Influence of Surface	
Similarity in Problem Solving as a Function	
of Knowledge	10
The Use of Past Examples In Categorization	
and Problem Solving	15
Problem Solving	15
Decision Making	19
Chapter 2	
Experiment 1	25
Methods	27
Results	31
Discussion	33

Chapter 3	
Experiment 2	38
Methods	40
Results and Discussion	43
Chapter 4	
General Discussion	46
The Influence of Similarity on Solving Elementary Probability Problems	47
The Influence of Similarity on Solving Pragmatic Inferential Reasoning Problems	52
References	56
Appendices	
Appendix A -- Problems Used In Experiment 1	62
Appendix B -- Problems Used In Experiment 2	78

List of Tables

**Table 1. Formula Correctness and Variable Instantiation
Scores as a Function of Training Group and Story
Line at Test 3 for Experiment 1.**

60

**Table 2. Mean Proportion of Problems Correct as a Function of
Story Line in Experiment 2.**

61

Acknowledgments

I am grateful to Dr. Masson for his help and comments on this work, and I extend thanks to the members of my dissertation committee for their comments. I would especially like to thank my brother for his comments, and my mom for her support and encouragement.

Dedication

To my Mother

CHAPTER ONE

Introduction

The purpose of this study was to determine whether individuals who understand the principle underlying a problem's solution will solve the problem by analogy to an earlier example as opposed to applying an abstract rule. Solving problems by analogy to past problems has typically been considered to be characteristic of novice problem solvers. Past research has shown that individuals who are learning a new domain often think back to an earlier example of which the current problem reminds them, and use the example and its solution to help them solve the current problem (Novick, 1988; Reed, Dempster, & Ettinger, 1985; Ross, 1984, 1987, 1989). This use of examples has been considered by some (e.g., Medin & Ortony, 1988; Polya, 1945; Wickelgren, 1974) to be a useful problem solving heuristic, in that individuals learning a new domain may not be able to apply the rule underlying a problem's solution because they do not understand the concepts used in the rule. As such, reliance on past examples as a guide to the current problem's solution is the only means available to solve the problem. Difficulties, however, emerge when problems are solved by analogy to earlier examples. For instance, as will be discussed in the subsequent sections of this dissertation, the current problem may not remind the individual of an earlier example, and in consequence the subject cannot draw an analogy between the new and an old problem. In addition, an inappropriate solution procedure may be applied to a new problem because the individual was reminded of an example that on the

surface appeared similar to the current problem, but in actuality employed a different solution procedure.

Past work (Bassok, 1990; Novick, 1988) has shown that individuals who understand the principle underlying a problem's solution do not require that an example of the current problem be available in memory in order for them to solve the current problem. For example, Novick (1988) found that when subjects were not reminded of an example of a similar problem, novices could not solve the current problem correctly. On the other hand, individuals who understood the principle underlying the problem's solution recognized the problem as employing the particular principle, and in turn applied the appropriate solution procedure. None of the work conducted to date, however, has investigated whether individuals who understand the principle underlying a problem's solution will nonetheless solve a current problem by analogy to a similar problem when such an example is readily available in memory.

The work reported in this dissertation indicates that individuals who understand the principle underlying a problem's solution are influenced by memory for past examples of problems. Specifically, individuals who understood the rule used to solve a problem nonetheless applied an inappropriate solution procedure when they were reminded of a problem that employed a solution procedure that was different from that used in the current problem. From a theoretical standpoint the finding that memory for prior examples continues to influence problem solving after the individual has developed an understanding of the problem's solution-relevant features is an interesting one. As will be discussed in this

dissertation, the findings of past research (Ross, 1984; Ross & Kennedy, 1990) have led to a better understanding of the occurrence and use of examples in problem solving. The work presented, however, provides a better understanding of when memory for past examples influences problem solving. That is, some researchers (Ross, 1984; Ross & Kennedy, 1990) have claimed that the use of examples is restricted to the early stages of learning when an understanding of the principle underlying the problem's solution has not yet been developed. Instead, the results of the present work suggest that memory for prior examples influences the behavior of both learners in a new domain and individuals with an understanding of the domain, but for different reasons. Specifically, novices in the problem domain, who do not understand the problem's underlying principle, rely on examples because they have no other means available to solve the problem. In contrast, individuals who possess knowledge that is relevant to the problem's solution may, when reminded of a past problem, use the example as a means of simplifying the retrieval and application of the problem's solution procedure.

As in past work that has investigated the influence of earlier examples on problem solving (Ross, 1984, 1987, 1989), subjects in Experiment 1 attempted to solve elementary probability word problems. Unlike past work, however, the influence of earlier examples was investigated using both participants who demonstrated an understanding of the principle underlying the problem's solution and those who did not understand how the problem was solved. By including subjects with different knowledge of the domain it was possible to determine whether only novices solve

problems by analogy to earlier examples, or whether the availability of previous examples influences problem solving regardless of differences in solution-relevant knowledge.

In Experiment 2, subjects attempted to solve problems that employed pragmatic inferential reasoning rules in their solution. The rationale for using this type of problem is that past work has shown that the rules are easily learned, and once learned the problems are solved through the application of the rules as opposed to solving the problems by analogy to earlier examples. As such, the conditions of this experiment provide a situation that allows for the investigation of whether individuals who consistently solve problems by using an abstract rule will nonetheless solve a problem by analogy to an earlier problem by default when such information is available in memory.

Reminders in Problem Solving

Work by Ross (1984, 1987; Ross & Kennedy, 1990) has demonstrated that individuals learning a new domain often retrieve examples of previously solved problem when attempting to solve a new problem. This retrieval has been termed a "reminding", and is characterized as being unintentional in the sense that the subject is neither asked to retrieve an example from memory, nor does the subject initially approach the task with the intention of searching memory for an example of the current problem. Once the example has been retrieved, however, the subject is aware of having previously experienced the problem during an earlier part of the experiment. Furthermore, reminders reflect the retrieval of information

from episodic memory in that variables that impair recall performance, such as interference, also decrease the occurrence of reminders.

Noticing Examples of Past Problems

In his work, Ross (1984) documented that reminders occur during the early stages of learning a skill and have predictable effects on performance. In one experiment, Ross trained subjects to perform text editing operations on a computer word processing system. Each operation could be performed using two different methods; for example a word could be moved by using a method that "appended" words, or a method that "inserted" into the text. Each subject was presented with two different texts (e.g., a poem and a restaurant review), and one of the editing methods was illustrated using the poem, whereas the restaurant review was used to illustrate the other method. Following training the subjects were asked to edit a new text (either a new poem or a new restaurant review), and instructed to "think aloud" or report their thoughts while attempting to edit the text. Ross reported that reminders occurred during problem solving in that the subjects' protocols indicated that they were recalling the previous training problem when trying to perform an operation (e.g., "the last time we had to move fresh strawberries", p. 383). The results tentatively suggest that the subjects were using the examples to guide them to the problem's solution, in that the subjects were more likely to successfully edit the text when the protocol indicated the occurrence of a reminding than when a reminding was not evident in the subject's statements.

The results of the experiment also indicated that the reminders were context specific in that the editing method retrieved was usually the one

illustrated in the training text that was similar to the test document. In turn, Ross suggested that the occurrence of a reminding depended upon the match between the current problem and the information stored in memory. Specifically, Ross (1984, 1987) claimed that whether a novice noticed an earlier example of the current problem depended upon the extent to which the earlier example and the test problem were superficially similar. Superficial or surface similarities refer to the features of a problem that are not relevant to the manner in which the problem is solved; examples of superficial features include the problem's story line, the names assigned to the variables, or the perceptual characteristics that are available when the problem is presented in the form of a diagram (e.g., slopes and pulleys in physics problems). The superficial features of a problem can be contrasted with structural features, or the abstract aspects of the problem that are relevant to the problem's solution (i.e., the principle underlying the problem's solution).

In a second experiment, Ross (1984) demonstrated that the occurrence of reminders is influenced by the superficial features shared by the current and the past problem, and more important, as suggested in the first experiment, that reminders influence problem solving performance. At study, Ross presented subjects, who had no previous training in probability or statistics (i.e., novices) with four elementary probability principles (i.e., combinations, permutations, waiting time, and selecting a specific choice at least once), along with a word problem that used the principle in its solution. The story line content of the problem was varied between study and test. For example, at study the story line content discussed a situation

in which readers of a film magazine selected baby photographs of actresses. Then, for each principle used at test, Ross presented problems in which the story line content was either: (a) different from that of a training problem that used the same principle (IBM scientists selecting computers); (b) similar to a training problem that used the same principle (newspaper readers selecting baby photographs of Canadian Prime Ministers; or (c) similar to a training problem that used a different, but related principle (e.g., a combination problem was substituted for a permutation problem). Ross reported that in comparison to the different story line problems, more problems were solved correctly when the test problems' story line was similar to a training problem that used the same principle (i.e., positive transfer effect). In contrast, more problems were solved incorrectly when the story line was similar to a problem that used a different, but related principle (i.e., negative transfer effect). The results indicated that the novices probed their memory for information that would help them solve the current problem, and in turn what was retrieved depended upon the match between the superficial features of the test problem and that of the training problem. In the case of novices, the similarity of the problem's superficial features influenced retrieval because the subjects did not possess an understanding of the problem's solution relevant features.

The Use of Reminders in Problem Solving

Ross's work on reminders has also provided a framework in which the question of what kind of information is retrieved from the example, and how the information is then applied to solve a problem can be addressed. Ross (1987) investigated which of two frameworks could best describe how

examples are used to solve problems. According to the principle-cuing view, when an example of an earlier problem is retrieved, the principle underlying the problem's solution is accessed and then directly applied to the problem. In contrast, the example-analogy view maintains that the earlier example and its solution are remembered, but the solution is applied by analogy to the current problem.

Ross (1987) demonstrated that the superficial similarities that the current and past problems share, cued not only the solution procedure that would be used, but also how the solution should be applied to the new problem. In this work, Ross trained subjects on four probability principles (mentioned in the previous section of this chapter). At test, the story lines of the problems were varied such that the content was similar to a study problem or new to the experiment. For each test problem, the subjects were provided with the appropriate solution formula and their task was to fill in the variables correctly with the numbers given in the test problem. In order to examine whether the current problems were solved by analogy to a past example, Ross investigated whether the details of the earlier problem would influence how the subjects instantiated the variables in the current problem's formula. The paradigm employed in the investigation was Gentner and Toupin's (1986) cross mapping technique. Here, the objects and the people in the word problem correspond to specific variables in the problem's solution formula. For example, in a permutations problem in which the subjects must choose an ordered subset of size r from a set a size n , the objects would correspond to n (the total number of objects) and the people would be assigned to r (the size of the subset); a cross mapping would

consist of the reversed assignment of objects and people to variables (i.e., the people would be assigned to n and the objects would correspond to r).

The object correspondence of the study and test problems in Ross's work were either the same (e.g., golfers (n) were randomly chosen to receive r prizes; fishermen (n) were randomly chosen to receive r prizes); or reversed (e.g., golfers (n) were randomly chosen to receive r prizes; prizes (n) were randomly chosen by r fishermen). In addition, the story line content of the study and test problems was either similar (as in the above example), or unrelated (e.g., IBM scientists (n) were randomly chosen to receive r computers; prizes (n) were randomly chosen by r fishermen). Ross predicted that if subjects were solving the current problem by analogy to an earlier example, then the similarity of the story line and the object correspondence between problems would be important. That is, the subjects would use the details of the example as a guide to how the variables should be instantiated into the formula. Specifically, the novices would look to the superficial similarities between the past and present problem for clues as to how the variables in the problem should be instantiated into the formula.

The results clearly supported the example-analogy view, in that when the current problem's story line reminded the subject of an earlier example, the subjects assigned the superficially similar objects to the same variable role that those objects played in the study examples. For instance, in comparison to when the current problem's story line did not remind the subject of a past example (i.e., new story line), the subjects were more likely to correctly instantiate the variables when they were reminded of a problem

in which the objects were assigned to the same variable roles. In contrast, subjects were more likely to instantiate the variables incorrectly (i.e., reverse the variables) when they were reminded of a superficially similar problem in which the variable to which the object corresponded was cross mapped. These results did not support the principle-cuing view because this view argues that reminders only serve to cue the appropriate abstract information, which is then directly applied to solve the problem. Instead, the results supported the view that superficial similarities between problems not only serve to retrieve an example of a similar problem from memory, but also influence how the information is then used to solve the problem.

The Differential Influence of Surface Similarity on Problem Solving as a Function of Knowledge

As previously mentioned, word problems possess both surface (superficial) features and structural features. Surface information is defined as the features of a problem that play no causal role in determining the solution to the problem; an example of which are the words that describe the statement of the problem. Structural information, on the other hand, refers to the features that are relevant to the problem's solution. This structural knowledge consists of an abstract understanding of the rule or principle that underlies a problem's solution. The rule is abstract in the sense that its application to a problem is not contingent upon forming an

analogy to an earlier example as a means of determining how the rule is employed. Instead, the rule is directly applied to the current problem.

Work that has investigated how individuals who understand a problem's structure in turn solve similar problems has shown that their focus on the problem is different from that of individuals who possess surface information only. Specifically, the latter group of individuals look to past superficially similar examples as a guide to the current problem's solution, whereas the former group disregard superficial information and instead apply the abstract rule. Evidence that individuals possessing qualitatively different representations of problems are differentially influenced by a problem's surface features was provided in work by Chi and her colleagues (Chi, Feltovich, & Glaser, 1981). In this study, novices and experts in the domain of physics were presented with a number of physics problems that used different laws of physics in their solution. In some cases, the problems shared the same underlying physics principle, but were superficially dissimilar (e.g., the statement of the problem used a picture of a slope, while another problem used a picture of a pulley). In other cases, the problems were superficially similar (e.g., all the problem statements used pulleys), but did not use the same principle of physics in their solution. The subject's task was to sort problems into categories of similar problems. Chi found that experts group the problems according to the underlying principles that the problems shared. Novices, on the other hand, classified the problems according to the superficial similarities that the problems shared. Similar research (Silver, 1979) that used mathematical word problems instead of physics problems, also found that

good mathematical problem solvers classified the problems according to the mathematical principle that the problems shared, whereas poor mathematical problem solvers sorted the problems on the basis of the similarities of the story lines used in the statement of the problems. Both of these studies indicate that individuals who possess an understanding of the principle underlying a problem's solution are not influenced by the superficial features of the problem.

Evidence for the notion that individuals who possess an abstract understanding of a problem's underlying principle focus on the problem's structural details, as opposed to its superficial details, when solving new problems has been presented in transfer studies (e.g., Novick, 1988; Reed, 1987). In these studies, the subjects reviewed a problem and its solution during a training phase, and later were asked to solve a structurally similar, yet superficially different problem (i.e., isomorphic problems). For example, at study, Novick (1988) presented subjects with three mathematical word problems. The solution to one of the problems required that the subjects find the Lowest Common Multiple (LCM) of the first three divisors, and then add a constant to the multiples of the LCM. The statement of the problems was presented using a story line that discussed finding vegetables in a garden. At test, subjects were presented with a problem in which the solution again required the subject to find the LCM. The story line of the test problem, however, discussed the task of finding marching band members in a high school. Novick reported that poor mathematics problem solvers did not apply the solution used in the training problem to the isomorphic transfer problem. She did, however, note that

good mathematics problem solvers did show a transfer of knowledge to the superficially dissimilar transfer problem, thus indicating that the good problem solvers did not rely on the availability of a superficially similar example of the current problem as a guide to the current problem's solution.

As noted by Ross (1984) the superficial similarities that exist between problems tend to determine whether novices will show positive transfer to structurally similar problems (e.g., Gentner & Landers, 1985; Pirolli & Anderson, 1985; Reed, Ettinger, & Dempster, 1985). For example, Holyoak and Koh (1987) presented subjects with a training problem in which the story line content discussed how low-intensity X-ray beams emanating from different directions could be used to converge upon and safely destroy a tumor (as opposed to one strong ray which would destroy the tissue surrounding the tumor). The subjects were then presented with the problem of how laser beams could be used to destroy a light bulb enclosed in a fragile case. In this instance, the subjects noticed the surface similarity between the transfer and training problem, and used the knowledge acquired at study to solve the later problem. In contrast, Gick and Holyoak (1983) failed to find a transfer of knowledge of the solution to the tumor problem when they used an isomorphic problem, in which soldiers were to be dispatched to a fortress, such that one large group of men could not be captured at once. Only subjects who had developed an abstract understanding of the tumor problem, said that the solution to the fortress problem was to divide the troops into smaller forces and converge upon the fortress.

The results of the transfer studies indicate that only individuals who have not developed an understanding of a problem's underlying structure rely on past examples of similar problems as a guide to a new problem's solution. In their study examining how examples are used in the construction of an abstract representation of a principle, Ross and Kennedy (1990) noted that subjects were less likely to solve problems by analogy to earlier examples once they attained an understanding of the problem's structure. As in the previous transfer work discussed, the story lines of the problems used in Ross and Kennedy's study were unique. Therefore, although it may be the case that individuals with structural knowledge solve problems by directly applying an abstract rule, they may solve a problem by analogy to an earlier example when the problem's story line reminds them of a previous example of the current problem.

Specifically, the objective of the work reported in this dissertation was to determine whether individuals who understand a problem's underlying structure will be influenced by the heuristic (used by novices) that problems that are superficially similar are solved using the same solution procedure. Medin and Ortony (1988) have suggested that reliance on superficial similarities to access a problem's solution relevant features is, in most cases, well justified in that there is often a non-random relationship between how something looks and its underlying structure. For example, things that have wings can fly, things that are round can roll, and things that are made of wood can burn. However, unlike nature, the surface features of mathematics word problems are usually not causally linked to the problem's underlying structure. As demonstrated by Ross (1984),

learners operate on the assumption that superficially similar problems possess the same underlying structure. That is, as previously discussed, when learners were reminded of a superficially similar yet structurally different example, they tended to apply the inappropriate formula to the current problem.

Reliance on superficial similarities is often the only solution strategy possible for novices who cannot abstract the structural similarities among problems, but realize that superficial similarities often signal related structure. None of the work conducted to date, however, has evaluated whether this heuristic may be used by default whenever a person is faced with a problem that is superficially similar to a previous example as a means of simplifying the retrieval and application of a principle. The motivation for the hypothesis that both novices and experts use this heuristic (but for different reasons) comes from work in the areas of categorization and decision making that have demonstrated that the influence of memory for example continues long after people have learned an abstract rule that can be applied to produce a correct response. That literature is reviewed in the following section.

The Use of Past Examples in Categorization and Decision Making

Categorization

Past work in categorization (Estes, 1986; Homa, Sterling, & Trepnel, 1981; Malt, 1989; Medin & Schaffer, 1978; Nosofsky, 1986) has demonstrated that categorization performance is based on the retrieval of specific past

instances, as opposed to the retrieval of abstract information. For example, Medin and Schaffer (1978) have argued that individuals initially learn some examples of a concept and then classify a new instance on the basis of how similar it is to a particular example of the concept. In essence, the new example reminds the individual of a similar old example, and classification is based on the assumption that similar items belong to the same category. The exemplar view of categorization is similar to Ross's (1984; Ross & Kennedy, 1990) notion of reminders, in that the similarities that exist between a current problem and a past problem determine whether a reminding will occur, and learners assume that superficially similar problems are similar types of problems (e.g., Ross, 1984, Exp 2).

It may be the case that episodic memory influences both problem solving and categorization in a similar manner. Whereas Ross has not investigated whether reminders continue to influence problem solving after the individual has developed an understanding of the problem's underlying structure, recent work by Allen and Brooks (1991) has indicated that memory for prior examples influences categorization performance even when the individual had learned a classification rule that when applied produces a correct response.

In their paradigm, Allen and Brooks presented subjects with drawings of imaginary animals that belonged to either the category "diggers" or "builders." Each animal varied on five binary dimensions (long or short legs, long or short neck, angular or curved body, two or six legs; spots present or absent) and category assignment was determined by the simple rule that an exemplar had to possess at least two of three features that

defined a category. For example, a "builder" was defined by the rule that category members had to possess long legs, an angular body, and spots-- animals that did not possess two of these features would be deemed a "digger."

During training, subjects learned the classification rule. They were then shown exemplars of each category and asked to classify each exemplar into its correct category by applying the classification rule. The subjects' knowledge of the classification rule was demonstrated in that they were able to classify correctly the exemplars into their appropriate category.

At test, subjects were asked to categorize items that were seen during training, and a matched set of new items that were visually similar to the old items. Furthermore, the new items were divided into "positive" and "negative" matches. A positive match was an item that belonged to the same category as the training item to which it had been matched. A negative match was an item that although visually similar to a particular training item belonged to a different category when the categorization rule was applied. For example, a new animal was visually similar to a past exemplar of a digger (e.g., both had a curved body and long legs), but because it was spotted and had long legs the rule would dictate that the new item be categorized as a builder. When subjects erred by calling the new item a digger, classification performance was assumed to be based on the comparison of the new item to a previously encountered category exemplar as opposed to the application of the simple rule.

The higher error rate for negative matches than for positive matches found by Allen and Brooks indicated that subjects were categorizing the

new exemplars on the basis of their similarity to a past example rather than according to the categorization rule. This result demonstrated that the use of examples extended beyond the early phases of learning as subjects who were able to correctly categorize items according to a simple rule during training later categorized new items on the basis of their similarity to a previously seen category exemplar as opposed to applying the known rule. Allen and Brooks concluded that when a salient exemplar of a category is readily available in memory, subjects will, by default, make category judgments on the basis of the item's similarity to previously seen category exemplars.

When they used materials that had greater ecological validity than imaginary animals, Brooks, Norman, and Allen (1991) similarly found that the availability of a visually similar category exemplar in memory influenced how both doctors with an average of 15 years of experience in family practice and doctors in their first year of residency classified photographs of skin disorders into their appropriate diagnostic categories. At study, subjects were shown photographs of skin disorders along with the diagnostic categories to which they belonged. Two weeks later, subjects were presented with new photographs of skin disorders. Half of the items were visually similar to the category exemplars seen at study, and half were visually dissimilar. The subject's task was to select from a number of choices the particular diagnostic class of skin disorders to which the pattern of skin ailment belonged; if diagnostic classification was made on the basis of applying a rule that defined the features that are characteristic of a particular class of skin disorders, then both the visually similar and

visually dissimilar exemplars should have been classified equally often into a particular category because all the exemplars were equally similar to the category prototype. Brooks et al. however found that disorders visually similar to the example were more likely to be assigned to the diagnostic category than visually dissimilar disorders that belong to the same diagnostic class of skin disorders. Category judgments therefore appeared to be affected by the new item's similarity to a category exemplar. Moreover, this pattern emerged for both groups of doctors, suggesting that the use of examples continues long past the early stages of learning a domain.

These two categorization studies used stimuli that require extensive perceptual processing and in turn one might argue that superficially similar examples may influence performance only when such materials are used. Work on decision making suggests that this is not the case.

Decision Making

Prior to the 1970's, human decision making was conceptualized in terms of a normative theory that maintained that when people make statistical decisions they rely upon readily accessible intuitions that correspond directly to major principles of probability and statistical theory (Edwards, 1968; Peterson & Beach, 1967). The work of Kahneman and Tversky (1972, 1973; Tversky & Kahneman, 1974, 1977, 1981, 1982a, 1982b), which now dominates decision making research, demonstrated that the normative theory could not be used to conceptualize human decision making because people very often violate principles of probability and statistics. For example, the conjunctive rule states that a conjunction, A

and B, is never more probable than either of its constituents. In an example provided by Tversky and Kahneman (1982a; 1983), a hypothetical woman, Linda, described as outspoken and concerned with issues of social justice, was judged more likely to be a feminist bank teller than a bank teller. According to the conjunctive rule, the decision was incorrect because if the subjects were considering true probabilities they should have been aware that the constituent bank teller was as least as probable as the conjunction feminist bank teller. Tversky and Kahneman argued that the judgment reflected the individual's reliance on similarities, as opposed to a principle, when making decisions. That is, in the case of the bank teller problem, Linda was more similar to the subject's concept of a feminist bank teller than to a bank teller.

The influence of similarity on decision making has also been demonstrated in work that did not involve probabilistic reasoning. Gilovich (1981) reported that superficial similarities between events can influence the predicted outcome of a new event. Moreover, similarity can influence decision making even when the individuals possess knowledge of the domain. Profiles of football players were presented to sportswriters and varsity football coaches, with the notion that the latter possessed more knowledge of the factors that impact on a football player's future success in professional sports. Subjects were given a profile of a college football player and a questionnaire in which they rated the player in terms of the probability of success in various aspects of professional football. Within each profile a comparison was made between the amateur athlete and a successful professional, but the factor that they had in common was

irrelevant in predicting the future success of the amateur football player. For instance, reference was made to the fact that both players came from the same hometown. Results indicated that subjects who had read a profile that made a comparison between the amateur and a successful professional football player were more likely to predict that the amateur would succeed in professional football, than subjects who had read a profile that did not mention a feature shared by both football players. The fact that both groups of subjects were influenced by the comparison made in the profile suggested that even when subjects possessed domain relevant knowledge they nonetheless reasoned by analogy when a superficially similar example was available.

In addition, Gilovich demonstrated that individuals with considerable knowledge in a domain might be biased to reason by analogy as a result of their memory for examples of previous events that could be associated with a current event. Gilovich presented political science students with profiles of foreign policy crises, and asked the subjects whether an interventionist or a "hands off" strategy should be adopted by the government. The two profiles of each crisis were the same with the exception that each contained a different feature that was irrelevant to the dynamics of the crisis, but that made one think of a similar real-world crisis. The critical feature made an incidental comparison to a real-world crisis in which either an interventionist (World War II) or hands off approach (Vietnam) proved to be the best strategy. For instance, mentioning that the ethnic minorities of the aggressor country were fleeing "by boxcars on freight trains" or by "small boats sailing up the coast" was intended to remind subjects of World War II

and Vietnam respectively. Results found that subjects reminded of World War II recommended more interventionist strategies than subjects in a neutral condition or one in which the superficial features reminded the subjects of Vietnam. In contrast, subjects reminded of Vietnam typically recommended a hands off strategy. This result draws attention to the paradox that a person with extensive knowledge of a domain may make an inappropriate decision as a result of drawing an analogy between a current problem and a readily available example of a similar event, even though features shared by the two cases actually are irrelevant to the decision.

The work conducted in the areas categorization and decision making demonstrates that memory for similar examples influences performance after the subject has attained an abstract understanding of the domain. It may also be the case that memory for past examples influence problem solving long after the person has developed an abstract understanding of the problem's underlying structure; particularly when the subjects are faced with a problem that is superficially similar to a past problem. To date, none of the work on problem solving has examined whether subjects who possess an understanding of the principle underlying a problem's solution will be biased towards solving a problem by analogy to a superficially similar yet structurally different problem. Such an occurrence might best be demonstrated by using a paradigm in which being reminded of an earlier example hinders performance (i.e., negative transfer). Novick (1988) did use a negative transfer paradigm in which subjects of low and high math ability (whom she referred to as novices and experts respectively) were required to solve a problem that was superficially

similar yet structurally dissimilar to an earlier problem. She found that both experts and novices demonstrated negative transfer in that for both groups of subjects, more people incorrectly used the training problem's solution procedure than the appropriate solution procedure. However, the effect was larger for the novices. That is, more novices than experts used the incorrect solution procedure. Her experiment, however, was not designed to indicate whether individuals who possess an abstract understanding of the problem structure are influenced by the availability of a superficially similar problem. That is, she hypothesized that high math ability subjects would be better able to develop an abstract understanding of the principle underlying an algebra problem's solution, and in consequence could later use this knowledge when solving a new problem. Low math ability subjects on the other hand would not be able to abstract the principle underlying a problem's solution and therefore would have to solve problems by using examples of a similar problem. The results supported this claim; whereas 54% of experts were able to apply the correct principle to a new problem, only 22% of the novices could. Novick's finding that 46% of the experts used the incorrect solution procedure tells us that these subjects solved the new problem by analogy to an earlier example, but it does not tell us that the experts solved a problem by analogy in spite of possessing an abstract rule. Unlike Allen and Brooks (1991), Novick did not assess whether the subjects had indeed acquired the rule whose application would result in the correct response; the 46% of the subjects who did not apply the correct solution procedure may not have been able to abstract the principle underlying the problem's solution and therefore solved the problem by

analogy to an earlier example. The results of Novick's study may only tell us that people who are mathematically astute are more likely to understand the abstract mathematical structure of an algebra problem (and in turn apply this information) than people who are mathematically less gifted.

To address the question of whether examples are used to solve a problem even when one has a rule available requires a test for negative transfer with subjects who demonstrate an abstract understanding of the principle underlying the problem's solution. The following work will show that when faced with a problem that is superficially similar to a past problem, both individuals who do and do not understand the principle underlying the problem's solution will use the previous example as a guide to both the access and application of a principle to a new problem. In this study, subjects learned the principle underlying the problem's solution, and evidence that they understood the solution principle was presented. Subjects then tried to solve problems in which the story line was new to the experiment (i.e., new story line condition), and problems in which the story line was intended to remind them of an earlier problem that used a different principle in its solution (i.e., re-paired story line condition). A negative transfer effect was expressed as a lower proportion of problems solved correctly in the "re-paired" story line condition, in comparison to the "new" story line condition.

CHAPTER 2

Experiment 1

Elementary probability problems were used in this experiment, and rather than selecting two groups of people who differed in their knowledge of probability principles, people were recruited who had no previous training in probability and statistics. Subjects were differentially trained such that half would acquire an abstract understanding of the principles and half would not. To do this subjects were asked to either (a) read an explanation of the principle and the solution to a problem that used the principle or (b) read the information and then explain how the principle was used to solve the problem. Any errors in the subject's explanation were corrected by the experimenter. Subjects in the latter training group were predicted to develop an abstract understanding of the principle because they had to learn the goal structure of the problem in order to provide an accurate explanation of how the principle was used to solve the problem.

In order to demonstrate that only the subjects in the read and explain group, had developed an abstract understanding of the probability principles I constructed test problems that were superficially different, but structurally similar to the training problems. Gentner and Toupin's (1986) cross mapping technique was used to reverse the object correspondence of the test problems relative to that of the training problems. These problems were then assigned to one of three test phases. In the first test phase subjects were explicitly informed about which training problem the current problem resembled. It was predicted that the clue would provide the subjects in both training groups with an effective retrieval cue to the correct

principle. However, because the subjects from the "read" group were not expected to develop an abstract understanding of the problem's structure, we predicted that when instantiating the variables into the formula, the subjects would often reverse the object correspondence because they were solving the problem by analogy to the training problem. Conversely, subjects who had developed an abstract understanding of the principle would instantiate variables correctly because they were not using the training problems as a guide to solving the new problems.

At test 2, the problems were again superficially different from those encountered earlier, but this time the subjects were not provided with a clue as to which training problem the current problem was similar; the lack of any similarities in surface information precluded a basis for retrieval among subjects who had not acquired an abstract understanding of the probability principles used in the problem (Novick, 1988; Reed, 1987; Ross, 1984, 1989, Ross & Kennedy, 1990). In consequence, subjects from the "read" group would be less likely than subjects from the "read and explain" group to use the correct solution formula. Also, in comparison to subjects from the "read" group, I predicted that subjects from the "read and explain" group would be more likely to instantiate the variables correctly because they were not relying on past examples for clues as to how to assign the variables to their structural roles.

Finally, a negative transfer paradigm was used in the final test phase in order to show that subjects who demonstrated an abstract understanding of the probability principle nonetheless solved problems by analogy when an example of a superficially similar problem was available in memory. Here,

the story lines of the test problems were either new to the experiment or similar to that of a training problem that used a related but different principle. When the problem's story line was new to the experiment, I predicted that in comparison to subjects from the "read" group, subjects from the "read and explain" group would be more likely to both access the correct principle and then correctly instantiate the variables into the solution formula. However, for problems in the re-paired story line condition, I predicted that the story line would remind subjects of a superficially similar training problem, and by default the subjects from both groups would apply the principle employed in the similar-story line problem to the current problem. Similarly, subjects from both conditions were expected to instantiate the variables incorrectly because they were solving the problems by analogy to the training problems in which the object correspondence of the variables was the reverse of the current problem.

Method

Subjects. Participants were 64 University of Victoria undergraduates, none of whom had taken a course in probability and statistics. Subjects recruited from Introductory Psychology classes were awarded course bonus points in return for their participation. Volunteers recruited from upper level psychology courses participated without benefit of payment or course credit.

Design. The experiment was divided into four phases: training, test 1, test 2, and test 3. During the training phase the subjects studied four principles of elementary probability, an example of a problem that used the principle, and an explanation and worked-out solution to the problem. The

training method (read vs read and explain) was manipulated between subjects.

At tests 1, 2, and 3 the subjects attempted to solve a set of four problems, one problem per principle. The object correspondence of the variables in the test problems was the reverse of the object correspondence used at training. For test 1 problems only, the subjects were apprised of the training problem that used the same principle. For each problem at test 3, the content of each problem's story line (re-paired vs new) was manipulated within subjects.

Materials. The four elementary probability principles used in this study were permutations, combinations, waiting time, and obtaining a specific choice at least once. For counterbalancing purposes the first two principles and the second two principles were viewed as two pairs of related principles. At test 3, one pair was assigned to the re-paired story line condition, and the other pair was assigned to the new story line condition. Counterbalancing ensured that each pair of principles appeared equally often in each condition.

For each principle, four word problems were constructed, and for each problem the content of the story line was different. These 16 problems comprised Set 1. A second set of problems was constructed in which the story line of each problem was similar to that of a Set 1 problem that used a similar, but not the same principle (i.e., combinations and permutations). The two sets of problems were formed in order to construct the re-paired story line condition at test 3. For the Set 1 problems, the re-paired story line problems were the Set 2 problems that used a story line similar to a problem that employed a related, but different principle. Similarly, for the Set 2

problems, the re-paired story line problems were the Set 1 problems that used a story line similar to a problem that used a related but different principle. The two problems used in the re-paired story line condition at test 3 were the problems from the alternate problem set that shared a similar story line, but not the underlying principle, of one of the problem pairs that the subject saw at study (See Appendix A for an example).

Two versions of each problem set were constructed; in one version the problems' object correspondence consisted of people being randomly chosen to receive an object, and in the other version objects were chosen at random by people. Half the subjects were trained and tested on the Set 1 problems, and half received the Set 2 problems. Within these two groups, half the subjects received the version of the test problems in which objects were randomly selected at test 1, and the version of the problems in which people were randomly selected at tests 2 and 3. At test 1, the other subjects received the version of the test problems in which people were randomly selected, and at tests 2 and 3 they received the version of the problems in which objects were randomly selected. In this way, the object correspondence of the test problems was always the reverse of the correspondence used in the training problems. Counterbalancing ensured that each of the four problems per principle were assigned equally often to the training, test 1, test 2 and test 3 phases.

All the problems were presented in a booklet that was divided into five sections: a cover sheet requesting the subject's age, year of university, and mathematical background (i.e., courses taken, when, and the grade obtained); a study section that discussed each probability principle, followed

by an example of a problem that used the principle plus its solution; and the test 1, test 2, and test 3 problems. The problems within and between test sections were randomized with the constraint that two principles from the same pair would never be adjacent to one another. This constraint was included in order to prevent the use of very short-term memory. The back of each booklet page was filled with typed characters so that the markings on the next page would not be visible to the subject. Lastly, a booklet containing the solution formulas of the four principles was constructed, one principle per page, and in a different random order for each subject.

Procedure: Subjects were randomly assigned to one of the instructional conditions (read or read and explain). At training, subjects in the read condition were given 5 min to read the theory behind the principle, the example problem, and the explanation and solution procedure. Subjects in the read and explain condition were instructed to read the material at their own pace and when ready provide an explanation of the goal of the problem and the principle behind the problem's solution. Any erroneous information in the subject's explanation was corrected by the experimenter. Following training in all four principles, subjects in both groups received the test 1 problems and were instructed to try to solve the problem the best they could. At the bottom of the problem page, a clue was written in red ink, informing the subjects as to which training problem the current problem resembled by specifying the earlier problem's story line. Subjects were instructed to use the clue to help them solve the problem. For each problem, subjects were told to search the booklet of formulas for the formula needed to solve the problem, and then to write down the formula and

instantiate the variables in the problem into the formula variables. The subjects were not required to perform any calculations, and were allowed four minutes to solve each problem. Following each test 1 problem solution attempt, subjects were given the correct solution to the problem and asked to review the solution and their answer for 30 s. The same procedure was followed for each test 2 and test 3 problem, with the exception that the subjects were informed that neither clues nor solutions to the problems would be provided for these problems. Test 2 and test 3 were treated as one block of problems and the subjects were not informed that the first four and the last four problems comprised different test phases. The session lasted approximately 90 minutes for the read and explain condition, and 60 minutes for the the read condition.

Results

In all of the analyses reported in this study, the Type I error rate was set at the .05 level, and all planned comparisons were carried out using the Bonferroni correction for familywise error, which was set at the .05 level.

Scoring

Each problem was scored twice; once for whether the correct formula was used and a second time for the correctness of the variable instantiation. If the correct formula was used a score of 1 was assigned, and a score of 0 was allocated when the answer used an inappropriate formula. For each subject, the dependent variable at test 1 and test 2 was the average score taken across the four problems, and at test 3 the dependent variable was the average score taken across the two "re-paired" story line problems, and the two "new" story line problems. The variable instantiation performance was

conditionalized such that no score was obtained unless the correct formula had been selected. When scoring the variable instantiation performance, a score of 1 was given if the answer was exactly correct, and a score of 0 was assigned if the object correspondence was reversed. Partial credit was awarded if the object correspondence was correct, but an error in instantiating a variable not related to object correspondence occurred; 3/4 credit if one mistake occurred, 1/2 credit for two mistakes. The dependent variable was the average score obtained across the number of items that used the correct formula.

Performance

Test 1. For the formula correctness performance the proportions correct were .781 and .891 for the read and the read and explain conditions respectively, $t(62) = 1.961$, $p < 0.054$, $SM_e = 0.056$. For the variable instantiation performance, the mean score in the read and explain condition (.835) exceeded the mean score obtained in the read condition (.475), $t(62) = 5.506$, $p < 0.000$, $SM_e = 0.065$.

Test 2. The proportion of formulas correct was reliably higher for the read and explain condition than for the read condition, .953 versus .539, respectively, $t(62) = 7.964$, $p < 0.000$, $SM_e = 0.052$. Similarly, for the variable instantiation performance the average score was reliably higher for the read and explain condition (.940) than for the read condition (.595), $t(62) = 5.205$, $p < 0.000$, $SM_e = 0.072$.

Test 3: Formula Correctness. The test 3 data are displayed in Table 1. A separate mixed factor Analysis of Variance (ANOVA) performed on the formula correctness data revealed a reliable effect of training method,

$F(1, 62) = 17.98$, $MS_e = 0.133$, and story line, $F(1, 62) = 16.86$, $MS_e = 0.111$. The training method x story line interaction was not significant, $F < 1$.

Test 3: Variable Instantiation. The results of a separate mixed factor ANOVA on the variable instantiation scores showed a reliable effect of training method, $F(1, 53) = 33.28$, $MS_e = 0.201$, and a reliable effect of story line, $F(1, 53) = 5.04$, $MS_e = 0.133$. The training method by story line effect was not reliable, $F(1, 53) = 1.05$, $MS_e = 0.133$. A planned comparison indicated that the difference between the new and re-paired story line condition for the read and explain group was reliable, $F(1, 27) = 8.859$, $MS_e = 0.080$. The same comparison of read group's variable instantiation data was not reliable, $F < 1$.

Discussion

Evidence that most subjects in the "read and explain" group had acquired an abstract understanding of the probability principles was demonstrated in both the formula correctness performance and the variable instantiation performance. First, in comparison to the "read" group, subjects from the "read and explain" group more often employed the correct formula at test 2 and for the problems in the "new" story line condition at test 3; this result replicates Ross's (Ross & Kennedy, 1990) finding that subjects who had developed an abstract schema of a probability principle were more likely to apply the principle to superficially different problems than were subjects who did not possess knowledge of the problem's structure. Second, again consistent with Ross (1987), we found that subjects in the read only condition who did not understand a problem's structural properties: (a) failed to instantiate the variables correctly under

conditions in which surface similarities were not available to remind them of a previous example (test 2) and (b) reversed the object correspondence when they were cued to use a problem in which the object correspondence was cross mapped (i.e., test 1). In contrast, the subjects from the "read and explain" group were able to correctly instantiate the variables in conditions in which the surface information did not remind them of a previous example (i.e., test 2 and the "new" story line condition at test 3).

The prediction that subjects from both training groups would use the appropriate principle when cued at test 1 was not supported; instead subjects in the "read and explain" group selected the correct formula more often than the subject in the "read" group. This performance difference may reflect the "read and explain" group's better memory of the training problems. However, although subjects from the "read and explain" group may have remembered the training problems more accurately, they did not solve the test 1 problems by analogy to the training problems. That is, the object correspondence of the test 1 problems was the reverse of the training problems; if the subjects in the "read and explain" group had solved the test 1 problems by analogy to the training problem, they would have usually assigned the variables to the structural role in the formula incorrectly. Instead, subjects assigned the variables to their correct structural roles, indicating that the test problems were rarely ever solved by analogy to the training problems. The variable instantiation performance of subjects in the "read and explain" group suggests that they retrieved the principle from the example and then directly applied the solution procedure. The subjects in the "read" group did reverse the assignment of the variables,

thus indicating that they were solving the test problems by analogy to the training problems.

The results of test 3 showed that similarities in surface information influenced access to the solution procedure even when the individual demonstrated an abstract understanding of the principle underlying the problem's solution. Both groups demonstrated negative transfer when the problem's story line reminded them of a superficially similar but structurally different problem. That is, more problems were solved correctly when the problem's story line was new to the experiment as opposed to when the subjects were reminded of an inappropriate example. The lack of a training method by story line interaction for the formula correctness performance indicated that both groups of subjects were influenced equally by the belief that superficially similar problems share a similar structure.

In addition, being reminded of a superficially similar problem influenced how the variables were instantiated into the solution formula. The subjects from the "read and explain" condition were more likely to reverse the object correspondence when the problem's story line reminded them of an inappropriate training problem in which the objects were assigned to a different variable, than when the story line was new to the experiment. The lack of a reliable difference between the new and re-paired story line condition for the read group can be explained in terms of this group's poor level of performance in both the re-paired and the new story line condition. That is, the subjects in this group probably did not develop an abstract understanding of the problems' structure. As such, the subjects applied

the variables to the same structural roles to which they were assigned at training. Since the object correspondence of the training and the test problems was different, the subjects performed poorly in each of the test phases. In contrast, subjects in the "read and explain" condition had developed an understanding of the problem's structure, and in turn instantiated the variables correctly when the problem's story line was new to the experiment. However, when the subjects in the "read and explain" group were reminded of a training problem, they often incorrectly assigned each variable to the same role that it had occupied in the training problem.

With respect to Ross's (1987) work that investigated the utility of the principle-cueing and the example-analogy views (discussed in the introduction), the results of test 1 indicate that the principle-cueing view best conceptualizes the problem solving of individuals who possess an abstract understanding of the principle underlying the problem's solution, but the problem solving of learners who do not understand how a problem is solved is best conceptualized by the example-analogy view. However, this may hold only for conditions in which the surface features of the problem do not remind the individual of an earlier example. That is, the finding in test 3 that the subjects in the "read and explain" group reversed the problem's object correspondence mainly when reminded of a superficially similar problem indicates that individuals who understand a problem's underlying structure will nonetheless solve it by analogy to an example when reminded of that example. The example-analogy view best conceptualizes the problem solving of individuals who understand a problem's structure under conditions in which they are spontaneously reminded of a

superficially similar example, but not under conditions in which they are reminded of an earlier example by an explicit clue. This finding supports Medin and Ortony's (1988) claim that superficial similarity has a powerful influence on cognitive tasks.

The next experiment examined whether similarity influences the problem solving of individuals who consistently solve problems by employing an abstract rule.

CHAPTER 3

Experiment 2

The second experiment had subjects solve problems that require the instantiation of pragmatic inferential rules. These rules are characterized as being intuitive or already part of a person's pre-existing inferential repertoire, and in consequence do not need to be learned de novo; instead, training improves the individual's understanding of the rules (Nisbett, Fong, Lehman, & Chong, 1987). One group of pragmatic inferential rules that has been used in past work (Fong, Krantz, & Nisbett, 1986; Fong & Nisbett, 1991; Lehman, Lempert, & Nisbett, 1988; Nisbett, Fong, Lehman, & Chong, 1987), and that was used here, involves statistical reasoning using the law of large numbers. These rules include such concepts as: the sample value's degree of resemblance to the population value varies as a direct function of sample size and population variability; and the concept of regression toward the mean in which extreme values for an object or sample are less likely to be extreme when the object is re-examined or when a new sample is observed.

The use of problems that employ these statistical rules was fostered by past work that showed that subjects can easily acquire an abstract understanding of statistical concepts following only a brief training session (Fong, Krantz, & Nisbett, 1986; Fong & Nisbett, 1991). That is, during training, subjects studied examples of problems that employed particular statistical rules for their solution. The study problems were similar in story line in that the content of each example pertained to a specific domain

(e.g., sports). At test, subjects solved new problems that, like the study problems, required the instantiation of a particular statistical rule for their solution, but in which the story line content was specific to a different domain (e.g., occupational ability). Evidence that the subjects had acquired an abstract representation of the statistical concepts came in the form that the correct concepts were applied to the test problems regardless of whether the story line content was specific to the domain referred to in the study examples (Fong & Nisbett, 1991). This result contrasts with the work on "reminders" in which early in training subjects rely on the reinstatement of surface information in order to access the correct solution rule (Ross, 1984, 1987; 1989). As such, the use of pragmatic reasoning problems allows us to examine whether similarity influences performance under circumstances in which performance is usually based on abstract reasoning, as opposed to solving the problem by analogy to an earlier problem.

Experiment 2 employed a negative transfer paradigm. During training, subjects learned four statistical concepts that could be viewed as two pairs of related concepts. Each statistical rule was learned within the context of a problem, and the content of each problem's story line was different. At test, subjects tried to solve a set of problems, some of which possessed a story line pertaining to the same domain as that of the training problems, a story line from a domain different than the one experienced at training, and a re-paired story line. The story lines of the related concepts were re-paired such that the story line content reminded the subject of an earlier problem that was structurally different from the current problem. In this

experiment, evidence that the subject acquired an abstract representation of the rule was expressed as similar levels of statistical reasoning performance for the same domain story line problems and different domain story line problems. If similarity of problems continues to influence problem solving performance after the subject has acquired an abstract representation of the principle underlying the problem's solution, then in comparison to the different domain story line problems, subjects would be more likely to solve the re-paired story line problem incorrectly. This result would stand in contrast to past work that has shown that problems are solved by analogy only in the early stages of learning a particular domain (Gick & Holyoak, 1983; Ross & Kennedy, 1990)

Method

Subjects. Twenty eight University of Victoria undergraduates were recruited in the same manner as in Experiment 1. None of the participants had taken a course in probability and statistics, and each subject was tested individually.

Design. The experiment consisted of three phases: training, test 1, and test 2. The training domain (sports vs. ability testing) was manipulated between subjects. The problems' story line at test 1 and test 2 (same domain, different domain, and re-paired) was manipulated within subjects.

Materials and Procedure. The problems used in this experiment employed the following concepts. The first two concepts and the second two concepts can be regarded as pairs of related concepts.

Concept 1: problems require subjects to draw conclusions about a population from a single small sample.

Concept 2: problems involve conclusions drawn from a sample that was large, but also highly biased--arguments in these problems should be based on the sample bias, but not on the sample size.

Concept 3: problems pitted a small sample against a large sample.

Concept 4: problems are similar to those in Concept 3, except that the large sample was drawn from a population that was related to, although not identical to, the target population--arguments in these problems should be based on population differences, but not sample size differences.

During training, subjects also studied the concept of regression toward the mean (i.e., an outcome selected because of its extreme deviation is not likely to be maintained in a subsequent sample), and the concept of erroneously pitting a large sample against a plausible theory not founded on data.

Problems that employed these concepts acted as filler items during training and were not presented at test.

For both the sports and the ability testing domains there were two problems per concept. Of the 16 problems, 11 were taken from past work (Fong, Krantz, & Nisbett, 1986; Fong & Nisbett, 1991) and five were constructed for this study. The story line content of each problem was unique. The re-paired problems presented at test 2 were constructed by taking the story line of the training problems that used Concept 1 and Concept 3, and writing problems that used Concept 2 and Concept 4 respectively (See Appendix B for examples). Problems that employed

Concepts 2 and 4 were not used because it was not possible to construct a statistical reasoning problem in which sampling from an unrelated population is an irrelevant feature of the problem. That is, it would not be possible to construct a problem in which the correct solution employed Concept 1 (differences in sample size) in light of there being differences in the population from which the sample was obtained. Pilot testing indicated that the newly constructed re-paired problems were interpreted as problems that instantiated the concept intended by the experimenter.

The sports domain problems were randomly divided into two sets. One set was used at training and the other set was used at test 1 (the problems used at test 1 comprised the same domain story line condition). Counterbalancing ensured that each set was assigned equally often at training and at test 1. This resulted in two problems sets (Set S1 and Set S2). The same procedure was conducted with the ability testing problems resulting in Sets A1 and A2.

In order to construct the different domain story line condition at test 1, Sets S1 and A1, and Sets S2 and A2 were paired: the different domain problems for Set S1 were the training problems used in Set A1, and the different domain problems for Set S2 were the training problems used in Set A2. Similarly, the different domain problems in Set A1 were the training problems used in Set S1, and the different domain problems for Set A2 were the training problems used in Set S2. For each of the four problem sets, the problems in which the story line was similar to a training problem (i.e., re-paired story line problems) were presented at test 2.

The training and testing materials were arranged in booklet form. In the first part of the booklet subjects were given a one-page introduction to the Law of Large Numbers that explained that this law can be helpful in understanding and predicting events, especially under conditions of limited information. The second part of the booklet presented four critical problems and two filler problems, one problem per statistical rule, followed by an analysis of it in terms of the Law of Large Numbers. Subjects were instructed to read the problem and the analysis at their own pace, and when finished they were to explain how the Law of Large Numbers was used to solve the problem. Any errors in their explanation were corrected by the experimenter.

The last part to the booklet contained the eight test 1 problems and the two test 2 re-paired problems. The subjects were not informed that test 1 and test 2 were different phases. For each subject the training, test 1 and test 2 problems were arranged in unique random order with the restriction that problems sharing the same principle did not appear in consecutive order. The back of each problem page was filled with characters such that the subject could not see the markings on the next page. All ten test problems were open ended, and subjects were instructed to think carefully about each problem and then write down an answer that seemed sensible. Subjects were also instructed that they might find the Law of Large Numbers helpful in their attempts to solve the problems.

Results and Discussion

Each subject contributed six scores: the answers to the same domain story line problem that used Concept 2 and the same domain story line

problem that used Concept 4; the different domain story line problem that used Concept 2 and the different domain problem that used Concept 4, and the two re-paired story line problems presented at test 2. The average score of the two problems was calculated for each of the three story line conditions (e.g., same domain, different domain, re-paired). Since all of the subjects answered the problems using one of the statistical concepts, the answers were assigned a score of 1 if the correct statistical concept was employed or a score of 0 if the incorrect concept was used. The mean score obtained for each of the three of story line conditions can be found in Table 2.

The mean score obtained in each of the conditions was analyzed using a mixed factor ANOVA. The analysis revealed a reliable effect of story line, $F(2, 52) = 5.25$, $MS_e = 0.088$. Neither the effect of training domain nor the training domain by story line interaction were reliable, $F(1, 26) < 1$, and $F(2, 52) = 2.51$, respectively. Because the interaction approached significance, separate comparisons of the story line conditions were conducted for the sports training domain and the ability testing training domain conditions. For the sports training domain a comparison revealed that the difference between the same domain (.89) and the different domain (.82) story line was not reliable, $F < 1$. As such, the two groups were combined and the means (.86) were compared to the re-paired (.50) story line condition. The difference between the two conditions was reliable, $F(1,13) = 9.09$, $MS_e = 0.01$. For the ability testing training domain the difference between the same domain (.75) and the different domain (.85) story line problems was not reliable, $F < 1$. Therefore, the two groups were combined (.80) and

compared to the re-paired story line condition (.71). The difference between the two groups was not reliable, $F < 1$.

Consistent with the findings of previous studies that have investigated statistical reasoning, subjects were equally likely to apply the correct principle to the same domain problems and to the different domain problems, thus indicating that the subjects were not solving problems by analogy to similar examples. Although the subjects in the sports training domain apparently possessed an abstract understanding of the statistical concepts, they nonetheless applied the reasoning concept employed in an inappropriate training problem when the test problem's story line reminded them of that problem. The lack of a negative transfer effect for the "ability testing" training domain may be due to the problems used in the re-paired story line condition. That is, pilot testing indicated that the sports domain and the ability testing domain story line problems were interpreted as reflecting the concept of biased sampling. However, the information that was used to convey this concept in the ability testing domain problems may have made the concept of biased sampling more apparent than the information that was used to present the concept in the sports domain problems. As such, the subjects in the "ability testing" training domain condition may have been less likely to be influenced by the similarity between the problems used at study and the re-paired story line problems because the concept of biased sampling was much more unmistakable for these problems than was the case for the sports domain re-paired problems.

The result obtained in the sports domain training condition are striking because past work has shown that problems employing the Law of Large

numbers are not usually solved by analogy to earlier examples. For example, Fong and others (Fong, Krantz, & Nisbett, 1986) reported that subjects who learned the concepts in the context of an undergraduate statistics course spontaneously applied the concepts to a home telephone survey question conducted in the last week of the semester, and Lehman and his colleagues (Lehman, Lempert, & Nisbett, 1988) found that subjects who learned the Law of Large numbers in the context of an undergraduate course applied the statistical concepts to problems presented four years later the context of a reasoning test. Given the time span covered in these studies, it is unlikely that the subjects were solving the problems by analogy to the past examples. The results of this experiment, however, suggest that even under circumstances in which the subject typically can apply an abstract concept to solve a problem, the availability of a superficially similar example in memory may influence how the subject answers the problem; specifically, the subject may solve the problem by analogy to the earlier example.

CHAPTER 4

General Discussion

The Influence of Similarity on Solving Elementary Probability Problems

The results of Experiment 1 indicate that memory for past examples influences problem solving regardless of whether one possesses an abstract understanding of the problem's underlying structure. Specifically, it was found that surface similarity influenced which principle would be used and how the principle would be applied (i.e., variable instantiation). In the first experiment, subjects were differentially trained such that half would be likely to develop an abstract understanding of the probability principles and half would be less likely. Furthermore, the existence of this knowledge difference was demonstrated in test 1 and test 2 (cf. Novick, 1988). That is, in comparison to the "read" group, the subjects in the "read and explain" group were more likely to use the correct principle and to instantiate the variables correctly into the formula when a problem with a novel story line was tested. The use of the correct formula when solving isomorphic problems (i.e., superficially dissimilar but structurally similar problems) is a characteristic of individuals who possess an abstract understanding of the principle underlying the problem's solution. Past work (Novick, 1988; Reed, 1987; Ross & Kennedy, 1990) has found that individuals who do not understand a problem's underlying structure fail to show a transfer of knowledge to isomorphic problems. Instead, they tend to use the knowledge acquired at study when the current problem's story line reminds them of an earlier similar problem (Holyoak & Koh, 1987; Ross, 1984; Ross &

Kennedy, 1990). In contrast individuals who understand the problem's underlying structure show a transfer of knowledge to isomorphic problems (Bassok, 1990; Novick, 1988). For example, Bassok (1990) found that individuals who learned the concept of geometric progressions within the context of banking problems could later transfer this knowledge to problems in which the story line discussed issues relating to the domain of physics. The results obtained in test 1 and test 2 of the first experiment reported here are consistent with the results that would be expected if the subjects in the two groups differed in their knowledge of the problems' principles. That is, the subjects in the "read and explain" group applied the correct principle to the isomorphic problems more often than the subjects in the "read" group.

The variable instantiation performance provided further evidence that the subjects in the two groups differed in their knowledge of the problems' structure. Gentner and Toupin's (1986) cross mapping technique was used to change the structural roles that the variables played in the word problems at training and at test. Past work (Gentner & Toupin, 1986; Ross, 1987, 1989) has demonstrated that individuals who do not understand the principle underlying a problem's solution use an earlier example as a guide to assigning the word problem variables to the formula. For example, Ross (1987, 1989) found that when learners in a new domain were reminded of an example of a problem in which the object correspondence has been cross mapped, the subjects were more likely to instantiate the variables incorrectly (i.e., reverse the object correspondence) than when they were reminded of an example in which the object correspondence matched that of the current problem. Ross and Kennedy (1990), however,

reported that individuals who understood the problem's structure were less influenced by the cross mapping manipulation. In the present study, subjects in the "read" group were less likely than subjects in the "read and explain" group to instantiate the variables correctly in test 1 and test 2. The subjects in the "read" group tended to reverse the object correspondence when solving the test problems, thus indicating that they were using the training problems as a guide to solve the test problems.

The knowledge difference that existed between the two groups was perhaps best illustrated in the first test phase. In test 1, subjects in both groups were informed as to which training problem the current problem was similar, by a clue that specified the training problem's story line. The object correspondence of the test 1 problems was the reverse of correspondence used in the training problems. The subjects in the "read and explain" group were more likely to use the correct principle than the subjects in the "read" group. This difference may have resulted from the "read and explain" group's better memory of the training problems. Although the subjects in the "read and explain" group may have better remembered the training problems, the variable instantiation performance indicated that they did not solve the test 1 problems by analogy to the earlier example. That is, in comparison to the "read" group, the "read and explain" group was less likely to incorrectly assign the object in the word problem to the same structural role that it had played in the training problem. This result indicates that the subjects in the "read" group were solving the current problem by analogy to the example. In contrast, whereas the explicit clue may have reminded the subjects of an earlier

example, the subjects in the "read and explain" group apparently solved the problem by directly applying the principle, because this group of subjects instantiated the variables correctly more often than the "read" group.

Despite their understanding of the probability principles, the subjects in the "read and explain" group demonstrated a negative transfer effect in test 3. That is, in comparison to when the problem's story line was new to the experiment, more problems were solved incorrectly when the story line reminded the subjects of a training problem that used a related but different principle in its solution. The finding that subjects were more likely to use an inappropriate principle when reminded of a superficially similar but structurally different problem is consistent with Novick's (1988) finding that both subjects of high and low math ability showed a negative transfer effect when the problem's story line reminded them of an inappropriate example. A curious result obtained by Novick was that fewer of the high math ability subjects chose the inappropriate solution procedure when training induced the subjects to compare the structural features between problems such that they would notice the structural differences between problems. In our study, the subjects in the "read and explain" group did notice the structural differences between problems because training highlighted the differences. The subjects' ability to use the appropriate principle in test 1 and test 2 affirmed that they noticed (and understood) the structural differences between problems. Regardless of this knowledge, the magnitude of the negative transfer effect for the "read" and the "read and explain" group was similar. A possible explanation for the difference between studies may have arisen from the difference in the problems used in this study and in

Novick's work. Novick used algebra problems that required the expenditure of a considerable amount of effort before a solution was reached, in that the median time required by her high math ability subjects was 8.7 minutes (15 minutes for the low math ability subjects). The current study, on the other hand, used elementary probability problems that could be solved with considerably less effort. It may be the case that the influence of similarity on problem solving is diminished under circumstances in which the problem's solution cannot be quickly accessed and easily applied. It is important to note that the negative transfer effect in Novick's study was diminished, but not eliminated. Along with the results of the current study, there is evidence to suggest that similarity exerts a powerful influence on the retrieval of a solution procedure regardless of the subject's understanding of the problem's structure. Moreover, the results of the variable instantiation performance in test 3 indicate that the availability of a similar example in memory influences not only which solution procedure is accessed, but also how the problem is then solved.

The "read" group's variable instantiation performance in test 1 is consistent with Ross's (1987) finding that learners in a new domain instantiate the variables into the formula by analogy to an earlier example, thus supporting the example-analogy view of the use of examples in problem solving. The results of the current study revealed that the principle-cueing view best conceptualizes the problem solving of individuals who understand a problem's structure, but only under conditions in which the person is reminded of the example by an external cue, as opposed to the surface similarities that exist between the current and the past problem.

That is, the principle-cueing view maintains that a principle is retrieved from an example, and then the principle is directly applied to the current problem. This view best conceptualizes the "read and explain" group's test 1 variable instantiation performance, in that when the subjects were reminded of the training problem by the external cue, they nonetheless instantiated the variables correctly. This finding indicated that the "read and explain" subjects were not using the training problem as a guide to the variables' structural roles. In test 3, however, the subjects were more likely to instantiate the variables incorrectly (i.e., reverse the object correspondence) in the "re-paired" story line condition than in the "new" story line condition. Interestingly, this result suggests that subjects who understand the principle underlying a problem's solution (as evidenced in test 1 and test 2), nonetheless solved the problem by analogy to an earlier example when the problem's story line reminded them of a training problem.

The Influence of Similarity on Solving Pragmatic Inferential Reasoning Problems

The results obtained in Experiment 2 also demonstrated that the use of examples occurs after the individual has attained an abstract understanding of the principle underlying a problem's solution. In order to provide a situation in which problems are usually solved through the retrieval of an abstract rule, as opposed to by analogy to an earlier example, I used problems that required the application of pragmatic inferential reasoning concepts for their solution. Past work (Fong, Krantz, & Nisbett, 1986; Fong & Nisbett, 1991; Lehman, Lempert, & Nisbett, 1988; Nisbett,

Fong, Lehman, & Chong, 1987) has shown that the retrieval of pragmatic reasoning concepts is not context specific (context specificity is the hallmark of reliance on memory for previous examples). That is, the concepts, learned in one specific context, generalize to different test contexts. An example of this generalizability is learning concepts in the context of problems in which the story lines involved sports, and testing using problems in which the story lines pertained to ability testing (Fong & Nisbett, 1991). Furthermore, there is evidence that the application of pragmatic reasoning concepts does not depend on memory for examples. Fong and Nisbett (1991) reported that following a 2-week delay between training and test, subjects reliably applied the statistical reasoning concepts to new problems, regardless of the fact that the subjects could not recall the examples that were used to convey the reasoning concept at training.

Similar to the results obtained by Fong and Nisbett (1991), subjects in Experiment 2 were able to apply the correct reasoning concept equally well to the same domain and the different domain problems presented in test 1. This result indicated that the subjects used abstract reasoning concepts to solve the problems, because when problems are solved by analogy to past examples, subjects do not typically show a significant transfer of knowledge to superficially dissimilar problems (Gick & Holyoak, 1983; Novick, 1988; Reed, Dempster, & Ettinger, 1985; Ross & Kennedy, 1990). In test 2, despite their abstract knowledge of the reasoning concepts, subjects in the sport domain training condition nonetheless applied an inappropriate reasoning

concept when they were reminded of a training problem that used a similar story line, but a different reasoning concept in its solution.

The results of Experiment 2 suggest that similarity has a powerful effect on problem solving. Problems that employ pragmatic reasoning concepts usually are solved through the application of an abstract reasoning concept (Fong, Krantz, & Nisbett, 1986; Fong & Nisbett, 1991; Lehman, Lempert, & Nisbett, 1988; Nisbett, Fong, Lehman, & Chong, 1987). As previously mentioned, however, the relative ease with which the concepts could be applied to the problems may have influenced the impact of memory for past examples on problem solving. Perhaps if the problems had required more effort in order to reach a solution, or if a longer retention interval was used, similarity may not have had such a strong effect on performance.

The use of problems that employed pragmatic inferential reasoning schemas in their solution provided a situation in which we could investigate how memory for prior examples influences the problem solving of individuals who consistently solve problems by employing an abstract rule. The results suggested that subjects who do solve problems by applying an abstract rule are influenced by superficial similarities between problems. The finding by Chi, Feltovich, and Glaser (1981) that experts disregard surface information when sorting problems into categories of similar problems could be attributed to the nature of the task. Specifically, the word problems were visible to the subject, and therefore classification did not require that the subject search through memory for problems that shared a similar principle. Perhaps the superficial aspects of the problems would have influenced card sorting if the expert subjects had to retrieve the

word problems from long term memory, and then classify them. That is, the problem solving heuristic by which superficially similar problems are solved using the same solution procedure may more likely influence performance when the use of the heuristic can simplify the retrieval of a solution procedure from memory.

Future research should investigate whether the influence of memory for prior examples on problem solving is diminished when problems that do not require the expenditure of a considerable amount of effort for their solution are used, whether the use of prior examples is a conscious phenomenon or reflective of a process of which the subject is unaware, and whether the effect is maintained over longer retention intervals.

I conclude with a qualification of the results of this study by noting that the effects of superficial similarity have in the past been found to depend not only on whether the study and test problems have a similar story line, but also on the confusability of the principles. Specifically, Ross (1987) reported that when he used problems that employed the principles of permutations and combinations (both involve the selection of a subset of items from a large set), the superficial similarities that existed between the study and test problems influenced performance. However, when the principles were made distinctive by substituting the very different principle of conditional probability for the combinations problems, the effect of similar story lines disappeared. It may be that heuristic reliance on similarity will negatively influence problem solving only when, as in this study, the principles could be easily confused with one another.

References

- Allen, S. W., & Brooks, L. R. (1991). Specializing in the operation of a rule. Journal of Experimental Psychology: General, 120, 3-19.
- Bassok, M. (1990). Transfer of domain-specific problem solving procedures. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 522-533.
- Brooks, L. R., Norman G. R., Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. Journal of Experimental Psychology: General, 120, 278-288.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems in experts and novices. Cognitive Science, 5, 121-152.
- Edwards, W. (1968). Conservatism is human information processing. In B. Kleinmuntz (Ed.), Formal Representation of Human Judgment (pp. 17-52). New York: Wiley.
- Estes, W. K. (1986). Memory storage and retrieval processes in category learning. Journal of Experimental Psychology: General, 115, 155-175.
- Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical reasoning on thinking about everyday problems. Cognitive Psychology, 18, 253-292.
- Fong, G. T., & Nisbett, R. E. (1991). Immediate and delayed transfer in training effects in statistical reasoning. Journal of Experimental Psychology: General, 120, 34-45.
- Gentner, D., & Landers, R. (1985). Analogical reminding: A good match is hard to find. In Proceedings of the International Conference on Systems, Man, and Cybernetics, (pp. 607-613). New York: IEEE.
- Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. Cognitive Science, 10, 277-300.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical reasoning. Cognitive Psychology, 15, 1-38.
- Gilovich, T. (1981). See the past in the present: The effect of associations to familiar events on judgments and decisions. Journal of Personality and Social Psychology, 40, 797-808.

- Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. Memory and Cognition, 15, 332-340.
- Homa, D., Sterling, S., & Trepnel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. Journal of Experimental Psychology: Learning, Memory, and Cognition, 7, 418-439.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80, 237-251.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47, 263-291.
- Lehman, D. R., & Nisbett, R. E. A longitudinal study of the effects of undergraduate training on reasoning. Developmental Psychology, 26, 952-960.
- Lehman, D. R., Lempert, R. O., & Nisbett, R. E. (1988). The effects of graduate training on reasoning in everyday-life events. American Psychologist, 43, 431-442.
- Malt, B. C. (1989). An on-line investigation of prototype and exemplar strategies in classification. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 539-555.
- Medin, D. L., & A. Ortony, (1988). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.) Similarity and Analogical Reasoning. Cambridge: Cambridge University Press.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification. Psychological Review, 85, 207-238.
- Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. Science, 30, 625-631.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. Journal of Experimental Psychology: General, 115, 39-57.
- Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 510-520.
- Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. Psychological Review, 68, 29-46.

- Pirolli, P. L., & Anderson, J. R. (1985). The role of learning from examples in the acquisition of recursive programming skills. Canadian Journal of Psychology, 39, 240-272.
- Polya, G. (1945). How To Solve It. Princeton, NJ: Princeton University Press
- Reed, S. K. (1987). A structure-mapping model for word problems. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 124-139.
- Reed, S. K., Dempster, A., & Ettinger, M. (1985). Usefulness of analogous solutions for solving algebra word problems. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 106-125.
- Ross, B. H. (1984). Reminders and their effects in learning a cognitive skill. Cognitive Psychology, 16, 371-416.
- Ross, B. H. (1987). This is like that: The use of earlier problems in the separation of similarity effects. Journal of Experimental Psychology: Learning, Memory and Cognition, 13, 629-639.
- Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on access and use of earlier problems. Journal of Experimental Psychology: Learning, Memory and Cognition, 15, 456-468.
- Ross, B. H., & Kennedy, P. (1990). Generalizing from the use of earlier examples in problem solving. Journal of Experimental Psychology: Learning, Memory and Cognition, 16, 42-55.
- Silver, E. A. (1979). Student perceptions of relatedness among mathematical verbal problems. Journal of Research in Mathematics Education, 10, 195-210.
- Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. Science, 185, 1124-1131.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the rationality of choices. Science, 211, 453-458.
- Tversky, A., & Kahneman, D. (1982a). Judgements of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgement under uncertainty: Heuristics and biases, (pp. 84-98). Cambridge University Press.

Tversky, A., & Kahneman, D. (1982b). Evidential in past of base rates
In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgement under
uncertainty: Heuristics and biases, (pp. 153-162). Cambridge University
Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive
reasoning: The conjection fallacy in probability judgments.
Psychological Review, 90, 293-315.

Wickelgren, W. A. (1974). How To Solve Problems: Elements of a Theory of
Problem Solving. San Francisco: Freeman.

Table 1
Formula Correctness and Variable Instantiation Scores as a Function of
Training Group and Story Line at Test 3 for Experiment 1.

<u>Training group</u>	<u>Story line</u>			
	<u>New</u>		<u>Re-paired</u>	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
<u>Formula Correct</u>				
<u>Read and Explain</u>	<u>.828</u>	<u>.273</u>	<u>.531</u>	<u>.358</u>
<u>Read</u>	<u>.500</u>	<u>.359</u>	<u>.313</u>	<u>.342</u>
<u>Variable Instantiation</u>				
<u>Read and Explain</u>	<u>.914</u>	<u>.257</u>	<u>.703</u>	<u>.455</u>
<u>Read</u>	<u>.391</u>	<u>.452</u>	<u>.313</u>	<u>.435</u>

Table 2

Mean Proportion of Problems Correct as a Function of Story Line in Experiment 2

	Story line					
	Same domain		Different Domain		Re-Paired	
<u>Training domain</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Sports	.893	.212	.821	.249	.500	.439
Ability Testing	.750	.325	.857	.234	.714	.323

Appendix A: Problems Used in Experiment 1:

The first two pairs of problems were used in Set 1 and the last two pairs of problems were used in Set 2. The letters "a" and "b" signify the object correspondence of the problem.

The Set 1 and Set 2 problems on each page can be viewed as problems with similar story lines, but solved using a related, but different principle.

Permutations a

The supply department at IBM has to make sure that scientists get computers. Today, they have 11 IBM computers and 8 IBM scientists requesting computers. On the basis of seniority (starting with the oldest), 3 scientists randomly choose their computer. What is the probability that the 3 scientists who have been at IBM the longest will get the lowest, second lowest, and third lowest serial numbers, respectively, on their computers?

Permutations b

The supply department at IBM has 11 IBM computers and 8 IBM scientists requesting computers. Today, the 3 IBM computer that are in operating condition are assigned on the basis of their serial numbers (lowest assigned first) and the scientist to receive an operating computer is randomly chosen. What is the probability that the 3 oldest scientists in progression of their age (oldest first) will get the lowest, second lowest, and third lowest serial numbers, respectively, on their computers?

a (Combinations)

The design department at GMC has just promoted 13 engineers on the basis of their seniority with the company. If there are 11 computers and 3 of the engineers request a computer this year, what is the probability that the 3 oldest engineers at GMC will randomly select the 3 computers with the lowest memory capabilities?

b (Combinations)

The design department at GMC has just promoted 13 engineers on the basis of their seniority with the company. If there are 11 computers and the 3 engineers who will receive these computers are selected at random, what is the probability that the oldest engineers will receive a computer?

Permutations a

West Hill High School has a vocational car mechanics class in which students repair cars. One day there are 12 students and 15 cars requiring repairs. Six students based on their rank on the last mechanics exam (the highest grade going first and so on) randomly choose a car to work on. If the cars vary in the extent of their damage, what is the probability that the 6 cars are worked on by the 6 students whose abilities are equally matched with the car's damage (e.g., worst student works on least damaged car and so on)?

Permutations b

West Hill High School has a vocational car mechanics class in which students repair cars. One day there are 12 students and 15 cars requiring repairs. Six cars are assigned according to the ranked severity of their damages (worst damaged going first), and the student who will work on a car is randomly chosen. What is probability that the 6 cars are worked on by the 6 students whose abilities are equally matched with the car's damage (e.g., worst student works on least damaged car and so on)?

a (Combinations)

The Newton Technical College offers a mechanics course in which students learn to fix motorcycles, and the students write their final exam in alphabetical order. This year there are 14 students and 17 motorcycles needing repairs. If at the end of the course 6 students randomly choose a motorcycle, what is the probability that the 6 poorest functioning motorcycles are worked on by the 6 students with the highest grades on the last exam?

b (Combinations)

The Newton Technical College offers a mechanics course in which students learn to fix motorcycles and the students write their final exam in alphabetical order. This year there are 14 students and 17 motorcycles needing repairs. If at the end of the course 6 motorcycles are assigned and the student who will fix a motorcycle is randomly chosen what is the probability that 6 of the poorly functioning motorcycles are worked on by the 6 students with the highest grades on the last exam?

Permutations a

This year the Mahomet Marathon attracts 31 runners, exactly the same 31 as the year before. If this year 15 of the runners are randomly given a number from 1 to 17 to wear, what is the probability that the first 8 runners each wear the same number as the year before?

Permutations b

This year the Mahomet Marathon attracts 31 runners, exactly the same 31 as the year before? If this year the number worn by the runners range from 1 to 21 and a runner is randomly selected to wear one of these numbers, what is the probability that the first 8 numbers are worn by the same runner as the year before?

a (Combination)

The Nanomat Triathlon attracts the same 37 entrants who participated in last year's Triathlon. If this year 15 of the entrants are randomly given a number between 1 and 17 to wear (of which 15 were numbers used the year before), what is the probability that last year's numbers are worn by this year's entrants?

b (Combinations)

The Nanomat Triathlon attracted 37 entrants this year (15 of whom participated in last year's Triathlon). If this year only the numbers ranging between 1 and 15 are issued and the entrants who will receive these numbers are randomly selected, what is the probability that last year's entrants wear this year's numbers?

Permutations a

Allan of "Al's Stereo and TV" has decided to specialize in stereos and therefore raffles off the 4 TVs in stock (a 120, 140, 180 and 240 centimetre screen TV) to any of the 5 customers in the store provided that they have bought over \$800 in stereo equipment. If 3 customers meet this criterion and going from the highest spender down to the lowest spender they randomly choose a TV what is the probability that the customer who spent the most money at Al's gets the largest TV, the one who spent the least get the smallest TV, and the customer who spent an intermediate amount gets the 140 cm TV.

Permutations b

Allan of "Al's Stereo and TV" has decided to specialize in stereos and therefore raffles off 3 of the 4 TVs in stock (the 120, 140, and 240 centimetre screen TVs) to any of the 5 customers who bought over \$100 in stereo equipment. If the TVs are raffled off on the basis of their size (starting with the smallest screen and working up) and the customers who will receive a TV are selected at random, what is the probability that the customer who spent \$200 received the smallest TV, the customer who spent \$101 received the largest TV, and the customer who spent \$150 received the 140 centimetre screen TV.

a (Combinations)

Ellen owns a record store and on the store's 1st anniversary she offers 3 CD players (worth \$500, \$600 & \$700) to any of the 5 customers in the store given that they have bought more than 9 CDs (equaling approx. \$180 in merchandise). If 2 customers have a receipt for the purchase of more than 9 CDs and select at random a CD player what is the probability that the most expensive CD players are selected?

b (Combinations)

Ellen owns a record store and on the store's 1st anniversary she offers 3 CD players (worth \$500, \$600 & \$700) to any of the 5 customers in the store given that they have bought more than 9 CDs (equaling approx. \$180 in merchandise). If 4 customers have a receipt for the purchase of more than 9 CDs and Ellen selects at random who gets a CD player what is the probability that the people who spent the most money will get a CD player?

Combinations a

A local kennel owner has hired 7 dog trainers who vary in their years of experience to teach dog obedience classes. On registration day the 3 most experienced trainers are working, and 5 dogs are brought to the kennel by their owners. Of these dogs 2 were older (7-10 years) and 3 were younger (2-4 years). If at this time each trainer randomly selects a dog to work with what is the probability that the 3 most experienced trainers work with the 3 younger dogs?

Combinations b

A local kennel owner has hired 7 dog trainers who vary in their years of experience to teach obedience classes. On registration day 5 dogs are brought to the kennel by their owners and of these dogs 2 were old (7-10 years) and 3 were young (2-4 years). If at this time a trainer is randomly selected to work with a dog what is the probability that the 5 most experienced trainers work with one of the 5 dogs?

a (permutation)

A local stable owner has hired 8 horse trainers who vary in their years of experience (2-9 years) to train horses for pleasure riding. One day on the basis of experience (most experienced goes first) 4 trainers randomly pick one of 11 horses to work with. Four of the horses were old (6-9 years) and 4 horses were young (2-5 years). What is the probability that the 4 trainers work with a horse whose age is the same as the rider's years of experience?

b (permutation)

A local stable owner has hired 8 horse trainers who vary in their years of experience (2-9 years) to train horses for pleasure riding. One day 8 horses arrive at the stable in sequence of their age (youngest first) and at this time a trainer is randomly selected to work with a horse. Four of the horses were old (6-9 years) and 4 horses were young (2-5 years). What is the probability that the 4 trainers work with a horse whose age is the same as the rider's years of experience?

Combinations a

An American airline has added more flights so 2 of their flight attendants must work an extra day a week (including weekends). The airline decides to enter each flight attendant into a lottery which will randomly choose the extra day that the attendant works. If every flight attendant is flexible as to which day they will work what is the probability that nobody works on a weekday.

Combinations b

An American airline has added more flights and, therefore, the flight attendants must work an extra 3 days a week (this included weekend days). The airline personnel manager decides that 8 flight attendants (of which 3 are newly hired) will be entered into a lottery which will determine, at random which attendant will work. If every flight attendant is flexible as to which day they will they will work what is the probability that the new flight attendants will have to rotate their work schedule?

A (permutation)

American Atlantic Railway is now providing rail service to more destinations. In order to do this 3 new customer service representatives are hired to work on 3 extra service days (this includes the week and weekend days). If at the time that the person is hired the personnel manager chooses at random the day that the person works, what is the probability that the first, second, and third representative hired work on Monday, Wednesday, and Saturday?

B (permutation)

American Atlantic Railway is now providing rail service to more destinations. In order to do this one customer service representative must work on Friday, and one must work on Saturday. The personnel manager places the names of the 3 representatives into a hat and draws at random the name of the person who works on Friday, and then draws the name of the person who works on Saturday. What is the probability that the oldest customer service representative works on Friday, and the youngest works on Saturday?

COMBINATIONS

Combination a

On the Labour Day weekend the Town Council sponsored a fish recipe fair at Bass Lake. This year 10 amateur cooks (of which 3 belong to the same family) are in attendance because there are 7 big ticket prizes (3 of which are cars) that can be won. If the 3 cooks who received a prize selected at random belong to the same family what is the probability this family will be 3 cars richer at the end of the day?

Combinations b

On the Labour Day weekend the Town Council sponsored a fish recipe fair at Bass Lake. This year 10 amateur cooks are in attendance because there are 7 big ticket prizes (3 of which are cars) and a cook will be selected at random to take one of these prizes home. If 3 of the cooks belong to the same family what is the probability that this family will be 3 cars richer at the end of the day?

a (permutation)

On the Labour Day Weekend the Chamber of Commerce sponsors a fishing competition at Salmon Lake. This year 8 fishermen have entered because of the size of the cash prizes (1st, 2nd, and 3rd place receive \$50, \$40, & \$30 dollars respectively). This year only two fishermen catch a fish, and tie because they have caught the same size fish. As such, the prizes are awarded at random by letting the first fisherman to have caught a fish draw a prize from a fishing hat, and then letting the second fisherman to have caught a fish draw a prize. What is the probability that \$50 dollars and \$30 are the respective prizes?

b (permutation)

On the Labour Day Weekend the Chamber of Commerce sponsors a fishing competition at Salmon Lake. This year 8 fishermen have entered because of the size of each cash prize (1st, 2nd, and 3rd place receive \$50, \$40, & \$30 dollars respectively). This year 7 fishermen all catch the same size fish and the winners are therefore randomly determined by a draw. What is the probability that the the youngest fisherman receives \$40, the second oldest fisherman receives \$50, and the oldest fishermen receives \$30?

Combination a

During the day 10 pharmacists were preparing capsules. They had to prepare 50 capsules in the first 30 minutes of the hour, following which 8 pharmacists graded the quality of a capsule selected at random. At the end of the hour, only 22 capsules had been prepared of which 14 were antihistamines and 8 were antibiotics. What is the probability that 8 pharmacists inspected 8 antihistamines?

Combination b

During the day 10 pharmacists were preparing capsules. They had to prepare 50 capsules in the first 30 minutes of the hour, following which a capsule would be inspected by a pharmacist selected at random. At the end of the hour, only 22 capsules had been prepared of which 14 were antihistamines and 8 were antibiotics. What is the probability that 8 antihistamines were inspected by the 8 pharmacists who prepared most of the antibiotics?

a (permutation)

One day 9 biologists were preparing slides in their medical research lab. At the end of the day 5 slides were finished, of which 3 were blood samples (i.e., Blood Type A, B, and O) and 2 were tissue samples (i.e., skin and hair). Two biologists, one with ten years of experience and one with 6 months of experience, selected a slide at random for a quality control test with the understanding that the most experienced biologist picked first so that she could finish early and still have time to work on other things. What is the probability that the most experienced biologist inspects the complex skin sample, while the least experienced biologist inspects the routine Type O sample?

b (permutation)

One day 9 biologists who varied in their years of experience were preparing slides in their medical research lab. At the end of the day 5 slides were finished, of which 3 were blood samples (i.e., Blood Type A, B, and O) and 2 were tissue samples (i.e., skin and hair). The skin tissue sample must be inspected immediately because its components disintegrate rapidly, and therefore, it is always the first sample to be tested. If 2 biologists are selected randomly, one to inspect the hair sample and one to inspect the tissue sample, what is the probability that the most experienced biologist inspects the complicated skin sample and the least experienced biologist inspects the routine hair sample?

Geometric Probability Distribution a.

Tourism is a very competitive industry and people involved in it must come up with creative methods to entice vacationers to choose their establishment. To promote a hotel in Hawaii, the owners have imported thousands of oysters, 10% of them with pearls inside. When guests check in they are each given an oyster to open. What is the probability that the 5th guest receives the first pearl?

Geometric Probability Distribution a.

1b. Tourism is a very competitive industry and people involved in it must come up with creative methods to entice vacationers to choose their establishment. To promote a hotel in Hawaii, the owners have imported thousands of oysters, 10% of them without a pearl inside. When guests check in they are each given an oyster to open. What is the probability that the 5th guest receives the first pearl?

A (Binomial Probability Distribution)

The winter season is usually a slow time for the travel industry and tourist spot owners must come up with ingenious ways of getting a person to come to their place of business. A beach resort in Mexico tries to attract tourists in the winter season by giving away silver bracelets. When the guests arrive at the resort, 25% will be given a bracelet. What is the probability that at least one bracelet is given to the 5th guest.

B (Binomial Probability Distribution)

The winter season is usually a slow time for the travel industry and tourist spot owners must come up with ingenious ways of getting a person to come to their place of business. A beach resort in Mexico tries to attract tourists in the winter season by giving away silver bracelets. When the guests arrive at the resort, 25% will not be given a bracelet. What is the probability that at least one bracelet is given to the 5th guest.

Geometric Probability Distribution a.

This is the year that Americans are voting for their President and in many states for a Governor. In an election, 3 candidates are running for governor, a Democrat, Republican, and Independent. Only 20% of the people vote for the Independent candidate. A clerk opens the ballots one at a time and tallies the votes. What is the probability that the first vote for the Independent is the fourth vote the clerk sees?

Geometric Probability Distribution b.

This is the years that Americans are voting for their President and in many states for a Governor. In an election, 3 candidates are running for governor, a Democrat, Republican, and Independent. Only 20% of the people do not vote for the Independent candidate. A clerk opens the ballots one at a time and tallies the votes. What is the probability that the first vote for the Independent is the fourth vote the clerk sees?

A (Binomial Probability Distribution)

Every 4 years Americans choose a new President, and this year many states will choose a new Governor. On the local ballot are 10 offices. Each office has 3 candidates running for it, a Democrat, a Republican, and an Independent. If a voter randomly chooses a candidate for each office (with each candidate having an equal 33% chance of being selected), what is the probability that the voter would vote for one or more Independents?

B (Binomial Probability Distribution) Every 4 years Americans choose a new President, and this year many states will choose a new Governor. On the local ballot are 10 offices. Each office has 3 candidates running for it, a Democrat, a Republican, and an Independent. If a voter randomly chooses a candidate for each office (with each candidate having an equal 33% chance of not being selected), what is the probability that the voter would vote for one or more Independents?

Geometric Probability Distribution a.

A recent marketing study conducted by a prominent advertising agency has indicated that 60% of a population of consumers is reputed to prefer a particular brand, "Minto" toothpaste. If a group of consumers is interviewed, what is the probability that exactly 5 people must be interviewed to encounter the first consumer who prefers brand "Minto".

Geometric Probability Distribution b.

A recent marketing study conducted by a prominent advertising agency has indicated that 60% of a population of consumers is reputed to not prefer a particular brand, "Minto" toothpaste. If a group of consumers is interviewed, what is the probability that exactly 5 people must be interviewed to encounter the first consumer who prefers brand "Minto"?

A (Binomial Probability Distribution) A recent poll conducted in the "Living Section" of the local newspaper has indicated that 70% of people who shop at Shop 'n Save prefer to buy the brand of soap called "Cleano". If a group of shoppers are interviewed, what is the probability that out of 6 people, at least 1 of them prefer "Cleano" to any of the other brand sold at this store?

B (Binomial Probability Distribution) A recent poll conducted in the "Living Section" of the local newspaper has indicated that 70% of people who shop at Shop 'n Save do not prefer to buy the brand of soap called "Cleano". If a group of shoppers are interviewed, what is the probability that out of 6 people, at least 1 of them prefer "Cleano" to any of the other brand sold at this store?

Geometric Probability Distribution a

A local swimming pool has started to interview people for the job of first aid attendant. Suppose that 30% of the applicants for the first aid attendant position have advanced training in cardiac resuscitation. Applicants are interviewed sequentially and selected at random from the pool. What is the probability that the first applicant having advance training in cardiac resuscitation is found on the fifth interview?

Geometric Probability Distribution b

A local swimming pool has started to interview people for the job of first aid attendant. Suppose that 30% of the applicants for the first aid attendant position have no advanced training in cardiac resuscitation. Applicants are interviewed sequentially and selected at random from the pool. What is the probability that the first applicant having advance training in cardiac resuscitation is found on the fifth interview?

B (Binomial Probability Distribution)

The local elementary school is hiring a teacher's aid to work 3 days a week. Suppose that 60% of pre-school teacher's aid have taken an advanced course in emergency first aid. If teachers are interviewed one after the other and are randomly selected from a pool of applications, what is the probability that at minimum 1 of five teacher's aids interviewed has indeed taken this important course?

B (Binomial Probability Distribution)

The local elementary school is hiring a teacher's aid to work 3 days a week. Suppose that 60% of pre-school teacher's aid have not taken an advanced course in emergency first aid. If teachers are interviewed one after the other and are randomly selected from a pool of applications, what is the probability that at minimum 1 of five teacher's aids interviewed has indeed taken this important course?

Binomial Probability Distribution a

The Main Street Restaurant has many desserts, $\frac{1}{7}$ of which are pies. Due to the location $\frac{4}{9}$ of the customers are senior citizens and the manager gives each customer a free desert, chosen at random. What is the probability that at least one of the 4 senior citizens gets a pie.

Binomial Probability Distribution b

The Main Street Restaurant has many desserts, $\frac{1}{7}$ of which are pies. Due to the location $\frac{4}{9}$ of the customers are senior citizens and the manager gives a pie to a customer chosen at random. What is the probability that at least one of the 4 pies goes to a senior citizen.

A Geometric Probability Distribution

The Seventh Avenue Cafe serves many types of beverages, $\frac{2}{10}$ of which are the non-alcoholic drinks, Spritz. One night $\frac{4}{10}$ of the customers are children and the owner gives each customer a free drink, chosen at random. What is the probability that the first Spritz is given to the fifth customer?

B Geometric Probability Distribution

The Seventh Avenue Cafe serves many types of beverages, $\frac{2}{10}$ of which are the non-alcoholic drinks, Spritz. One night $\frac{4}{10}$ of the customers are children and the owner gives a Spritz to a customer chosen at random. What is the probability that 5 people pay for their drink before the first customer gets a free Spritz?

Binomial Probability Distribution a

Many utility companies have begun to promote energy conservation by offering discounts to consumers who keep their energy usage below certain established subsidy standards. Usually $\frac{4}{6}$ of these discounts are rebates. A recent EPA report notes that $\frac{7}{10}$ of the residents of the islands of St. Domain have reduced their electricity sufficiently to qualify for discounted rates. If 7 qualifying subscribers are given a discount chosen at random what is the probability that at minimum 1 rebate will be offered.

Binomial Probability Distribution b

Many utility companies have begun to promote energy conservation by offering discounts to consumers who keep their energy usage below certain established subsidy standards. Usually $\frac{4}{6}$ of these discounts are rebates. A recent EPA report notes that $\frac{7}{10}$ of the residents of the islands of St. Domain have reduced their electricity sufficiently to qualify for discounted rates. If 5 rebates are made available and residential subscribers are randomly selected from St. Domain what is the probability that at minimum 1 rebate will be offered?

A Geometric Probability Distribution

In order to compete with private companies, a provincial telephone company has begun to offer lower long distance rates to consumers who keep their phone calls below a certain time limit. Of these lower rates, $\frac{3}{10}$ partial refunds. A recent study notes that $\frac{6}{10}$ of residents of St. Jacques Islands have reduced the length of their calls enough to qualify for the cheaper rates. If 6 long-distance subscribers qualify and are randomly given a lower rate what is the probability that the 5th subscriber gets the first partial refund.

B Geometric Probability Distribution

In order to compete with private companies, a provincial telephone company has begun to offer lower long distance rates to consumers who keep their phone calls below a certain time limit. Of these lower rates, $\frac{3}{10}$ are partial refunds. A recent study notes that $\frac{6}{10}$ of residents of St. Jacques Islands have reduced the length of their calls enough to qualify for the cheaper rates. If 6 long-distance subscribers are randomly selected what is the probability that 5 people will get a movie voucher before the first reduced rate is awarded?

Binomial Probability Distribution a

Many employers are finding that some of the people they hire are not who and what they claim to be. Detecting job applicants who falsify their application information has spawned some new businesses: credential checking investigators of which 1/4 of the employees are men. "U.S. News and World Reports" (July 13, 1981) reported on this problem, and noted that one service in a two-month period found that 1/5 of all credentials examined were falsified. If a man randomly selects 5 applications what is the probability that at least one has been falsified?

Binomial Probability Distribution b

Many employers are finding that some of the people they hire are not who and what they claim to be. Detecting job applicants who falsify their application information has spawned some new businesses: credential checking investigators of which 1/4 of the employees are men. "U.S. News and World Reports" (July 13, 1981) reported on this problem, and noted that one service in a two-month period found that 1/5 of all credentials examined were falsified. If 5 applications had to be checked and an investigator was appointed at random what is the probability that at least one applications will be checked by a man.

A Geometric Probability Distribution

"Canada Newsworld" (July, 1993) has reported that 4/10 of people who work in the Foreign consulate have top ranking jobs. Many top ranking foreign officials do not have the credentials that they claim they have, and these officials could be a threat to the nation's interests. When asked about this, a government spokesman reported that 1/3 of the resumes sent to the consulate included false information. In a random check of 5 new applications what is the probability that the first false credentials are found on the last one.

B Geometric Probability Distribution

"Canada Newsworld" (July, 1993) has reported that 4/10 of people who work in the Foreign consulates have top ranking jobs. Many foreign officials do not have the credentials that they claim they have, and these officials could be a threat to the nation's interests. When asked about this, a government spokesman reported that 1/3 of the resumes sent to the consulate included false information. In a random check of 5 top ranking officials what is the probability that the first top ranking official is the 4th application investigated.

Binomial Probability Distribution a

A prominent group of medical researchers have claimed that experience has shown that $\frac{3}{4}$ of all persons afflicted with a certain illness don't recover. A drug company has developed new vaccines, of which $\frac{1}{4}$ only mask symptoms. Ten vaccines were selected at random and injected into patients; nine recovered shortly thereafter. Suppose that people do not recover from the disease. What is the probability that at least 1 of the patients has masked symptoms.

Binomial Probability Distribution a

A prominent group of medical researchers have claimed that experience has shown that $\frac{1}{4}$ of all persons afflicted with a certain illness don't recover. A drug company has developed new vaccines, of which $\frac{3}{4}$ only mask symptoms. Ten people with the illness were selected at random and injected with the vaccine; nine recovered shortly thereafter. Suppose that the vaccine was absolutely worthless, what is the probability that at least 1 of the 10 patients would recover.

A Geometric Probability Distribution

A recent conference on communicable diseases has reported that 75% of all persons who contract a certain syndrome do not recover. A pharmaceutical company has begun testing new drugs for this disease of which 25% can be safely taken by all people. In accord with CDFA rules, if 16 drugs are selected at random and tested what is the probability that 9 drugs will have to be tested before the first person shows no side effects.

A Geometric Probability Distribution

A recent conference on communicable diseases has reported that 75% of all persons who contract a certain syndrome do not recover. A pharmaceutical company has begun testing new drugs for this disease of which 25% can be safely taken by all people. In accord with CDFA rules, 16 people who have contracted this disease are selected at random and given the new drug. Of these people, what is the probability that 9 people will have to be tested before the first recovery from the disease occurs.

Appendix B: Problems Used In Experiment 2

Each problem is followed by an example of the re-paired problem.

Sports Domain

Concept 1 : Conclusions based on a single small sample

A talent scout for a professional basketball team attends two college games with the intention of observing carefully the talent and skill of a particular player. The player looks generally excellent. He repeatedly makes plays worthy of the best professional players. However, in one of the games, with his team behind by two points, the player is fouled while shooting, and has the opportunity to tie the game by making both free throws. The player misses both free throws, and then tries too hard for the rebound from the second one, committing a foul in the process. The other team then makes two free throws, for a four-point lead, and goes on to win by two points.

The scout reports that the player in question "has excellent skills, and should be recruited. He has a tendency to misplay under extreme pressure, but this will probably disappear with more experience and better coaching."

Comment on the thinking embodied in the scout's opinion that the player (a) "has excellent skills" and that the player (b) "has a tendency to misplay under extreme pressure." Does the thinking behind either conclusion have any weaknesses?

* Please consider the problem for a few moments. After you have considered the problem and analyzed it for a minute or two, turn the page for our analysis.

ANALYSIS

The scouting agent is trying to draw a conclusion about a certain population. We can think of the population as the college athlete's basketball plays. We can divide the population into two categories: those basketball plays that are equivalent to those of the best professional basketball players, and those basketball plays that are worse than those of the best professional basketball players. The scouting agent claimed that the college basketball player's game performance is similar to that of the best professional basketball players. The scouting agent is making a statement about the population, namely, that the college basketball player displayed a greater percentage of basketball plays that were equivalent to those of the best professional basketball players, than basketball plays that

were worse than those of the best professional basketball players. This conclusion was based on observing a sample of size 2; the 2 games that were attended by the college scout.

The Law of Large Numbers states that small samples are less reliable indicators of the population than are large samples. A sample of size 2 is so small that any conclusions drawn from it are very unreliable. Thus, the scouting agent's claim is quite unwarranted: he would need a larger sample to really know whether the college athlete's basketball plays are comparative to the performance of professional basketball players.

*** only appears on problems presented at training**

Re-paired: Concept 2 (biased sampling) but a Concept 1 story line

A talent scout for a professional basketball team was eager to recruit college basketball players who could compete well in games against the best of professional basketball players. The manager's assistant saw a player in two college games and thought he was not a particularly good competitor as he tended to lose control of his skills in each of the games. For example, he would try too hard for the rebound, and would commit a foul in the process. In turn, the other team would make the subsequent free throws and go on to win the game by a mere two points. The talent scout attended eight practice sessions with the intention of observing the talent and skill of one particular player. Unlike the assistant he notes that, "during the practice drills, the player showed tremendous control and agility as he dribbled the ball through the most challenging of patterns, and his shooting technique was the best of the team as he hardly ever missed a basket from a jump shot. In fact the players practicing around him usually watch in order to pick up some pointers on their own jump shot."

The talent scout phones NBA team headquarters and reports that the player in question "should be recruited. I saw his skills in the practices and I'm convinced that he would play superbly in our Western division games helping us knock the number 3 team from its current rank."

Comment on the thinking embodied in the scout's opinion that the player "should be recruited". Comment on his reasoning

Concept 1 : Conclusions based on a single small sample

A professional football team had for years maintained a strict policy on which college players they selected for the draft. They refused on principle to consider players from College Division 1-AA, the division composed of smaller colleges. Instead, they draft only players from Division 1-A, the division composed of the major universities.

One year later, a new scouting director was hired. Thinking that the team's policy passed up many quality players, the scouting director decided to try an experiment. That year, the team drafted two players from smaller colleges. Both of them made the team. In fact, both were judged after the first season as being above the team average in ability.

On the basis of this experiment, the scouting director wrote a report to the owner of the team stating that "...there seems to be no differences between players from Division 1-AA and those from Division 1-A. Players from smaller colleges seem to be just as motivated and talented as those at the larger universities."

What do you think of the scouting director's conclusion? Comment on his reasoning.

* Please consider the problem for a few moments. After you have considered the problem and analyzed it for a minute or two, turn the page for our analysis.

ANALYSIS

The scouting director is trying to draw a conclusion about a certain population. The population is composed of college football players from smaller colleges. We can divide the population into two categories: those who would perform above the professional team average, and those who would perform worse than the team average. The scouting director has made a statement about the population, namely, that there is no difference between players in the two divisions. Translating this, he is claiming that the percentage of quality players in this population is just as high as the percentage of quality players in another population--players from larger universities. This conclusion was based on observing a sample of size 2; the 2 players that were drafted from the smaller colleges.

The Law of Large Numbers states that small samples are less reliable indicators of the population than are large samples. A sample of size 2 is so small that any conclusions drawn from it are very unreliable. Thus, the scouting director's claim is quite unwarranted: he would need a larger sample to really know about the comparative abilities of players from the two divisions.

Re-paired: Concept 2 (biased sampling) but a Concept 1 story line

When selecting players during the draft a professional football team in the United States had made it clear that they refuse to recruit players from College Division 1-A, the division composed of smaller colleges. Instead they only draft players from Division 1-B, the division that consists of the major "Big Ten" universities.

This year, a new recruiting director was hired, and it was his opinion that they had missed out on many very excellent players because of the team's policy. That year, the team drafted twelve players from the larger colleges. Only five of these players made the team. In fact, only the four players were judged to be above the team's average in ability.

On the basis of this finding, the recruiting director wrote a report to the owner of the team stating "...that players from the smaller colleges are better players than the A-1 teams and should be considered for the team."

What do you think of the recruiting director's conclusion? Comment on his reasoning.

Concept 2 : Biased Sampling

A group of college coaches argued that their team players should receive some kind of guaranteed minimum income to help them pay for their expenses while attending college. They cited a recent study of several hundred professional athletes with a minimum income. "Nearly, 92% of those athletes" he said, "work very hard at their sport and wanted very much to play an instrumental role in taking their team to victory." He noted that this was contrary to the popular belief that athletes will lose their competitive drive once money becomes an incentive to participate in one's sport. "Thus a guaranteed income policy would result in an increase in the players' willingness to work hard and bring home the championship to the college." The coach also noted that donations to colleges that win collegiate championships increase 30%, and therefore, it would be in the best interest of the college to provide the players with a minimum income.

Do you agree with the coach? Why or Why not? Comment on his reasoning?

* Please consider this problem for a few minutes. After you have considered the problem and analyzed it for a minute or two, turn the page for our analysis.

ANALYSIS

The coaches are trying to draw a conclusion about a population. The population is defined as the college's football players. On the basis of observations taken from a sample of several hundred professional athletes, the college football coach draws the conclusion that a minimum guaranteed income will result in an increase in his football players' willingness to work hard and bring home the championship.

The Law of Large Numbers states that large samples are more reliable indicators of the population than are small samples. The Law of Large numbers does not, however, pertain to this situation. The college football coach wanted to make some conclusions about the effect of a guaranteed minimum income on the performance of his college football players. The population from which the sample of several hundred athletes was actually obtained, however, did not consist of college football players but rather professional athletes. In order to make any conclusions about a population on the basis of information obtained from a sample, the sample must be taken from the population that one wishes to draw conclusions about.

Concept 2 : Biased Sample

Martha was talking to a fellow rower. The fellow rower was new to the team and was willing to go to extra training sessions on Sunday afternoons. "I don't think that Sunday practice sessions will improve our rowing performance," Martha said. "I've been on the rowing team for six years now. I've found that when I missed several practice sessions I felt that I was not rowing as well as I could. When I saw that my rowing was not up to my usual standards, I just made sure I went to all the regular practice sessions and the problems disappeared. Really, I don't know why I missed so many sessions. I guess that I just find it hard to get motivated sometimes. Lack of motivation will hurt your rowing, but I disagree that extra practice sessions will improve our rowing at this stage of skill development."

Analyze Martha's reasoning. Do you think that she had good evidence for feeling she doesn't need the extra rowing practice?

* Please consider this problem for a few minutes. After you have considered the problem and analyzed it for a minute or two, turn the page for our analysis.

ANALYSIS

Martha was trying to draw a conclusion about a specific population, where the population was defined as Sunday afternoon practice sessions. The question of interest was whether these extra practice sessions improve rowing performance. The Law of Large Numbers states that large samples are more reliable indicators of the population than are small samples. The Law of Large numbers does not, however, pertain to this situation. In order to make any conclusions about a population on the basis of information obtained from a sample, the sample must be taken from the population that one wishes to draw conclusions about.

Martha drew the conclusion that extra rowing practices would not affect rowing skills on the basis of her many experiences where she missed regular rowing practices. In order to draw a conclusion regarding the influence of extra practice on rowing performance, Martha would have had to observe her rowing performance following her participation in extra practice sessions--the larger the sample of extra practice sessions, the better indicator of the true effect of extra practice on rowing she would have.

Concept 3 : Small sample pitted against a large sample

The coach of a professional hockey team was in a quandary. It was close to the beginning of the season and there was just one more position on his timetable that had to be filled--left wing. The problem was that it wasn't at all obvious which player should make the team. It came down to two players--Stephens and LaBreque. The coach had received reports on each of them from two scouts--one who had seen Stephens play five games in college, and one who had seen LaBreque for six games. In general, the scouting reports suggests that Stephens was a better skater, defender, and team player than LaBreque.

But the coach himself had not seen either of the two players, so during the afternoon practice session, he carefully watched Stephens and LaBreque. It turned out that LaBreque clearly outplayed Stephens in the practice session. LaBreque showed remarkable speed and agility on the ice, broke up two power plays, and displayed excellent teamwork. Stephens was rather lackluster. Although Stephens seemed like a competent and solid player, he didn't shine in any respect. The coach concluded, from watching the practice session, that the scouts were wrong. He concluded that LaBreque, not Stephens was the better skate, defender, and team player.

Which player do you think the coach should select for his team? Explain your answer.

* Please consider the problem for a few moments. After you have considered the problem and analyzed it for a minute or two, turn the page for our analysis.

ANALYSIS

The coach is trying to draw a conclusion about two populations. One of the populations is composed of Stephen's hockey playing ability and one population is composed of LaBreque's hockey playing ability. Each population can be divided into two categories: one category being solid hockey playing ability, and one category being less than competent hockey playing ability. The coach's conclusion that LaBreque was the better hockey player was based upon observing a sample of size 1; the practice session that he attended.

The Law of Large Numbers states that small samples are less reliable indicators of the population than are large samples. The coach had only seen a sample of size 1-- this sample is so small that any conclusions drawn from it are very unreliable. Thus, the coach's claim is quite unwarranted: he would need a larger sample to really know about the comparative abilities of the two players. The talent scout, on the other

hand, attended 5 of Stephen's games and 6 of LaBreque's games. Using the Law of Large numbers, the talent scout's claim the Stephen's is the better player is probably correct, in that the talent scout's larger sample of observations enabled him to make reliable conclusions about the hockey player's true ability.

Re-paired

Concept 4 (large sample but unrelated pop) and Concept 3 story line

The coach of a professional hockey team needed to recruit a hockey player. The beginning of the season was near and there was just one more position on his team that had to be filled - right wing - which needed a player with superior defensive skills. The problem was that it wasn't at all easy to determine which player should join the team. The team's talent scout had seen two players - Micheals and LaFleche - play in one of their games in the junior league. In general, the scouting reports claimed that Micheals was the better skater, defender, and team player.

The coach himself had not seen either of the two players, but was positive that Micheals should not be recruited because the coach of Micheal's team emphasized power plays to a greater extent than break away plays. The coach said to the talent scout, "I've seen over 20 games with junior players from teams that emphasize power plays and although the skills they're taught are adequate for the junior league, when they enter the major league their defensive skills, their skating ability, and their teamwork is substandard. Recruit LaFleche, his team's coach doesn't emphasize power plays. This way we won't have the team's manager fire us because Micheal's defensive skills can't match those of the other teams we play against."

Which hockey player should be recruited? Explain your answer.

Concept 3 : Small sample pitted against a large sample

Pat Clemons is the coach of the college gymnastics team. One of her jobs is selecting new gymnasts for the team. She says the following of her experience: "Every year high school coaches from all over the country send us gymnasts who they think have a lot of potential. These high school coaches report enthusiastically about these young people - a young woman who can do a brilliant tumbling run or a young man who does a combination of jumps that make you hold your breath. Unfortunately at the college team tryout, most of these young people turn out to be less talented than the coaches would have you believe. I have found that the best performances at the tryout have sometimes been given by gymnasts who are not rated very highly by their high school coach. You know, I think that high school coaches simply spend all their time with a few gymnasts and fail to recognize the talent and growth of the other gymnasts."

Comment on Pat's reasoning. Why do you suppose that Pat revises downward her opinion of gymnasts that high school coaches initially thought were brilliant?

* Please consider the problem for a few moments. After you have considered the problem and analyzed it for a minute or two, turn the page for our analysis.

ANALYSIS

Pat is trying to draw a conclusion about two populations. One of the populations is composed of gymnasts who are highly rated by their high school coach, and one population is composed of gymnasts who are not highly rated by their high school coach. Each population can be divided into two categories: one category being gymnasts who possess exceptional gymnastic skill and once category being gymnasts who possess only competent gymnastic skill. The coach's conclusion that the gymnasts who are not rated highly by their high school coaches actually possess the better gymnastics skills was based upon a sample of size 1; the gymnasts' performance at the 1 tryout. The college coach therefore made conclusions about the gymnast's skills on the basis of a very small sample of observations.

The Law of Large Numbers states that small samples are less reliable indicators of the population than are large samples. The coach had only seen a sample of size 1-- this sample is so small that any conclusions drawn from it are very unreliable. Thus, the coach's claim is quite unwarranted: she would need a larger sample of observations in order to really know about the comparative abilities of the two groups of gymnasts. The high school coaches, on the other hand, would have seen their gymnasts' perform on several occasions. Since the Law of Large Numbers states that

large samples are more reliable indicators of the population than are small samples, the high school gymnastics coach would be better able to reliably assess the gymnasts' true ability than the college gymnastics coach.

Re-paired

Concept 4 (large sample but unrelated pop) and Concept 3 story line

Olga Clark was hired by the coach of a college gymnastics team. Olga is a former champion and in recent years had been a judge at the World Championships and the Olympics. She was hired to help the coach select which college gymnasts would compete in the college nationals. Olga comments "the nationals are the hardest competition a college athlete will ever face. I think this stems from the quality of the competitors at the event, the high expectations of the judges, and the television coverage of the event doesn't exactly lessen the athlete's anxiety." During the tryout competition Olga judged each athlete on their ability to successfully accomplish the combination of compulsory tumbling runs that everyone competing in the nationals must perform in their routine in order to advance to a medal position. One gymnast performed the elements superbly, while a second gymnast had considerable difficulty. Olga had observed the second gymnast in eight previous training sessions prior to the tryouts and thought that she was quite good. Olga left the tryout and the next day phoned the coach with her decision regarding which gymnast should go to the nationals.

Did Olga choose the first or the second gymnast to compete at the nationals? Comment on the reasoning behind Olga's decision.

Concept 4 : Large sample taken from an different population

The coaches of the track and field team at South Eastern University keep records on the performances of all their team's athletes and relates this information to background information about the student athletes. Recently there was a debate about whether to recruit a particular student from Horace Maynard High School. The student's first place finishes at high school track meets were low -- that is, all student athletes who actually made it on South Eastern's track team had many first place finishes, while most rejected student athletes had few first place finishes. This particular student's ratings of future potential were good, but none of the high school track and field coaches who had rated the student were personally known to any of the coaches at South Eastern.

Coach A of South Eastern's coaching team argued against recruitment, pointing out that several students on track teams at small rural high school like Horace Maynard perform at levels substantially lower than students from larger, better funded schools. This argument was countered by a Coach B who noted that 2 years ago, South Eastern recruited a student from Horace Maynard who was now among the three highest ranked student athletes in the country.

Comment on the strength and weaknesses by these two coaches. What are their strength and their weaknesses?

* Please consider this problem for a few minutes. After you have considered the problem and analyzed it for a minute or two, turn the page for our analysis.

ANALYSIS

The coaches are trying to draw a conclusion about a certain population. We can define the population as the track and field athletes from Horace Maynard High School. The question asked is whether these student athletes have the talent to become high ranking college athletes. Coach B draws the conclusion that these students do have potential, in that a student from Horace Maynard is now among the highest ranked athletes in the country. This coach's conclusion is based on a sample of size 1; the 1 student from Horace Maynard High School. The conclusion drawn by the Coach A who stated the opposite view was based on the performance of a larger sample of students.

The Law of Large Numbers states that larger samples are more reliable indicators of the population than are smaller samples. Although the sample size observed by the Coach A was larger than Coach B's sample, we cannot say that Coach A's conclusion was a reliable indicator of the population because Coach A's sample was taken from a different population than the one of interest, namely Horace Maynard students. Coach A observed

students from "small rural schools *like* Horace Maynard", but did not observe several students from Horace Maynard High School. As such, Coach A's conclusions are meaningless when trying to assess the potential of student athletes from Horace Maynard High School's population.

Concept 4 : Large sample taken from a different population

It's the bottom of the ninth inning of an important baseball game, late in the season. There are two outs, and the home team is behind by one run. The manager of the home team has a decision to make. He has to choose which of two right-handed pinch hitters will bat next. Player A has a season batting average of .300; that is every time he comes to bat, he gets a hit 30% of the time. But Player A has never faced this particular pitcher. Player B, on the other hand, has a season batting average of .250 (every time he comes to bat, he gets a hit 25% of the time), but in the five times he's faced the pitcher this year, he's had two hits, giving him a batting average of .400 against this pitcher, that is a 40% chance of getting a hit against this pitcher.

Which player do you think the manager should use as the pinch hitter? Explain your reasoning.

* Please consider the problem for a few moments. After you have considered the problem and analyzed it for a minute or two, turn the page for our analysis.

ANALYSIS

For both Player A and Player B, the season batting average was based on a large sample of observations--a seasons worth of observations. The Law of Large Numbers states that large samples are more reliable indicators of the population than are small samples. Since the sample size is the same for Player A and Player B, one might assume that the manager should use Player A as the pinch hitter because he gets a hit 30% of the time while Player B only gets a hit 25% of the time. This, however, is not the correct answer.

The manager is not trying to draw a conclusion about Player A's and Player B's overall hitting ability. He is trying to draw a conclusion about each player's hitting ability when facing a particular pitcher. Therefore, one population consists of the number of hits Player A gets against this particular pitcher, and one population consists of the number of hits Player B gets against this particular pitcher. Since Player A has never faced this particular pitcher, we do not have a sample of hitting scores from which we could evaluate his ability to get a hit against the pitcher--the fact that Player A gets a hit 30% of the time does not tell us anything about his chances of getting a hit when facing this particular pitcher. In this problem, the manager should choose Player B to be the pinch hitter because he knows that player B has a 40% chance of getting a hit against this particular pitcher.

The following problems were only seen at training

Concept 5. regression towards the mean.

After the first two weeks of the major league baseball season, newspapers began to print the top ten batting averages. Typically, the leading batter after two weeks has an average of .450 (that is, every time he comes to bat, he gets a hit of 45% of the time). Yet no batter in major league history has ever had an average of as high as .450 at the end of the season.

Why do you think this is? Explain your answer.

* Please consider this problem for a few minutes. After you have considered the problem and analyzed it for a minute or two, turn the page for our analysis.

ANALYSIS

We can analyze this problem using the Law of Large Numbers in the following way. Every time a baseball player comes to bat, he either gets a hit or he doesn't. A time at bat can be considered a sample of size 1 from a population. But what is the population? It is the player's "true" hitting ability. For some players, their true ability is very high, corresponding to a batting average of .300 or better (that is, there is a 30% chance of getting a hit); for other players, their true ability is considerably lower. There are very few people whose true ability is higher than .300, and certainly no one whose true ability is as high as .450.

During the first two weeks of the baseball season, players have batted only about 30 times. This is quite a small sample. We know from the Law of Large Numbers that small samples are unreliable estimates of the population. It is entirely possible with a small sample that a player can hit well above his true average--he can hit a hot streak and bat .450 for two weeks.

Over the entire season, however, a player's batting average consists of a much larger sample--several hundred times at bat. We should expect the player's batting average at the end of a season to be a more reliable indicator of the player's true ability. Those who hit a hot streak over two weeks find it impossible to maintain that streak over an entire season. Thus, it is not surprising that after two weeks, there may be some players hitting around .450, but no player has ever batted that well over the entire season.

Concept 6: Theory not founded on data

A study recently conducted by researchers at North Central University examined the performance of 125 major league players in baseball, football, and basketball the year before they were married, and the year after they were married. From the results of the study, the researchers concluded that marriage did not have any effect on a player's performance: a player's performance the year before marriage did not differ at all, on average, from his performance the year after marriage.

Just after the study was reported, a sports writer from the local area criticized that study. He stated that it was obvious to him that marriage has a negative impact on a player's performance. According to the sports writer, marriage decreases a player's performance because the player has to assume many more responsibilities after marriage. It would be impossible, the writer said, for a major league player to maintain the kind of concentration necessary to perform at the same level. The writer stated that the conclusions of the study were wrong.

What do you think of the sport's writer's criticism of the study? Do you think it is well-founded? Explain your answer

* Please consider this problem for a few minutes. After you have considered the problem and analyzed it for a minute or two, turn the page for our analysis.

ANALYSIS

The goal of the survey was to assess whether marriage has a detrimental effect on the performance of professional athletes. The researchers took a sample of 125 athletes in order to test this hypothesis. The sample was drawn from the population of professional athletes. On the basis of the sample, the researcher made an inference about the population: that there were no difference in the performance of professional athletes after getting married.

The Law of Large Numbers states that as the sample size increases, the sample becomes a more reliable indicator of the population. A sample of 125 is quite a large sample. Apart from other considerations, we should trust the findings of the study.

What about the sports writer's theory? Although his theory is plausible, the opposite theory is equally plausible. One might expect the security and comfort of marriage to enhance the performance of a professional athlete. The point here is that a plausible theory that is unsupported by the data is not a good basis for judgments, especially when there is a large sample that

contradicts the theory. Thus, the sports writer's reasoning is not sound: It is more likely the case that marriage has no discernable general effect on the performance of professional athletes.

Ability Domain

Concept 1 : Conclusions based on a single small sample

Andrew was talking to a neighbor about his son Timmy, who was in the second grade. "Timmy loves reading and does very well in all the other things he's had in school so far, including counting, but I really don't think arithmetic is going to be his cup of tea. Last week we did several addition problems and he never showed any insight into the process; he just did them by rote. Then a few days ago we were working on some subtraction problems and he never got the idea of 'taking away.'"

Do you agree with Andrew's reasoning that "arithmetic isn't going to be Timmy's cup of tea"? Why or why not?

* Please consider the problem for a few moments. After you have considered the problem and analyzed it for a minute or two, turn the page for our analysis.

ANALYSIS

Andrew is trying to draw a conclusion about a certain population. Roughly speaking, we can define this population as Timmy's aptitude for arithmetic: Can Timmy solve arithmetic problems? Andrew's conclusions based on a sample of size 2; the 2 occasions on which Andrew observed Timmy working on arithmetic problems.

The Law of Large Numbers states that small samples are less reliable indicators of the population than are large samples. A sample of size 2 is so small that any conclusions drawn from it are very unreliable. Thus Andrew's reasoning is very poor: he would need a larger sample (more observations of Timmy working on arithmetic problems) in order to really find out about Timmy's aptitude for arithmetic.

Re-paired: Concept 2 (biased sampling) but a Concept 1 story line

Annette was discussing with a teacher her son Tommy who was in the fifth grade. "I think that Tommy should be put in a more specialized math programme because I think he will have difficulty with high school math. You see, Tommy has excelled in many of the topics he's encountered in school, but he's an avid reader. I read a study that used over 500 high school students that showed that high school students who liked to read are not motivated to do well in mathematics. Since Tommy is an avid reader now, I think a specialized mathematics program will motivate him to do well in high school mathematics later.

Do you agree with Annette? Comment on her reasoning.

Concept 1 : Conclusions based on a single small sample

Judy is a personnel director for a California based airline, in charge of hiring flight attendants. Recently, the number of passengers that speak only Spanish has risen, so Judy decided to screen applicants for their ability to speak Spanish. During the personal interview, applicants are asked to translate five English words into Spanish. If an applicant is able to correctly translate 4 of the 5 words, he or she will be given higher consideration in the hiring process, based on their knowledge of Spanish.

What do you think of this screening procedure? Is it a reliable indicator of the applicant's knowledge of Spanish?

* Please consider the problem for a few moments. After you have considered the problem and analyzed it for a minute or two, turn the page for our analysis.

ANALYSIS

The goal of Judy's screening procedure is to draw a conclusion about a certain population. Roughly speaking, we can define this population as a particular applicant's knowledge of Spanish: Is the applicant able to speak Spanish? Conclusions that are made from this screening procedure are based on a sample of size 5--the 5 words that the applicant is asked to translate into Spanish.

The Law of Large Numbers states that small samples are less reliable indicators of the population than are large samples. A sample of size 5 is so small that any conclusions drawn from it are very unreliable. Thus Judy's screening procedure is very poor: she would need a larger sample (longer Spanish test) in order to really find out about the applicant's knowledge of Spanish.

Re-paired: Concept 2 (biased sampling) but a Concept 1 story line

Jane was in charge of hiring flight attendants for a large Canadian airline. Unlike the other airlines it was this airline's policy to hire only applicants who could converse in both English and French since the majority of the airline's flights were to destinations within Canada. In order to screen applicants for their ability to converse in both French and English, Jane administered 300 items from the French Language Fluency Test (FLFT) to each applicant. Jane used this test because the vocabulary scores typically correlated quite well with a person's ability to converse fluently. If an applicant is able to correctly translate 280 of the 300 words,

he or she will be considered to be fluent in English and given higher consideration in the hiring process. If the applicant scores lower than 280 items correct they will not be considered for the job.

Comment on the validity of Jane's screening procedure.

Concept 2 : Biased Sampling

A young psychologist was talking to a college dean: "verbal ability is not very helpful for understanding the sciences. I suggest that students be admitted into the Bachelor of Science programme regardless of their score on the verbal section of the College Entrance Examination (CEE)." The dean retorted, "I'm not so sure that this is indeed the case. I just finished reading a study done with 600 chemistry majors at three university campuses showing that those students with low verbal ability (a CEE score of below 500) do more poorly than students with high verbal ability (a CEE score of above 600). I think that we should restrict admittance into the Bachelor of Science programme only to those students with high verbal ability."

Do you agree with the college dean? Why or Why not? Comment on his reasoning?

* Please consider the problem for a few moments. After you have considered the problem and analyzed it for a minute or two, turn the page for our analysis.

The college dean is trying to draw a conclusion about a population. The population is defined as the college science students. On the basis of observations taken from a sample of six hundred chemistry majors, the college dean draws the conclusion that only students with high verbal ability will excel in the Bachelor of Science programme

The Law of Large Numbers states that large samples are more reliable indicators of the population than are small samples. The Law of Large numbers does not, however, pertain to this situation. In order to make any conclusions about a population on the basis of information obtained from a sample, the sample must be taken from the population that one wishes to draw conclusions about. The college dean wanted to draw a conclusion about college science majors on the basis of the performance of 600 chemistry students. The sciences consist of more than just chemistry, and therefore, the sample does not represent the entire population of college science students. The Law of Large Numbers states that the college dean would probably have been correct in concluding that verbal ability is related to performance in the area of chemistry.

Concept 2 : Biased Sampling

Professor Smith reviewed with his students the material that would be covered on the first exam of Experimental Psychology 200. Smith was more than happy to provide a review session because he had attended a large university where the class population usually numbered over 150 students, and this severely limited the feasibility of a question and answer session. At the end of the class Smith commented, " When I was an undergraduate about 89% of the people in my class who failed the first exam went on to fail the course. The last thing any of you folks need is an F on your transcript. If you fail the first exam, I suggest that you get me to sign a course withdrawal form before the deadline comes and goes. Remember if you don't drop the class by the deadline, you're obligated to finish the course." Since Smith's class could hold 25 students, waiting list students might be admitted to this important course if some registered students decided to drop out.

What do you think of Professor's Smith's advice? Explain your answer.

* Please consider this problem for a few minutes. After you have considered the problem and analyzed it for a minute or two, turn the page for our analysis.

ANALYSIS

Professor Smith is trying to make a conclusion about a population. The population is the students who are in Smith's Experimental Psychology class, and the question of interest is whether the students who failed the first exam will also fail the course. The Law of Large Numbers states that large samples are more reliable indicators of the population than are small samples. The Law of Large numbers does not, however, pertain to this situation. In order to make any conclusions about a population on the basis of information obtained from a sample, the sample must be taken from the population that one wishes to draw conclusions about.

Smith drew a conclusion about the possibility of a student failing the present Experimental Psychology course on the basis of observing the scores obtained by several students in the undergraduate course that he attended as a student. In order to draw any conclusion about the probability of failing Smith's course, Smith would have had to observe the performance of a sample of students who were taking his Experimental Psychology course.

Concept 3 : Small sample pitted against a large sample

Dorothy was the head of a summer camp. The week before camp was to start, the bookkeeper quit. A new one had to be hired immediately. There was a small accounting school in the area which Dorothy called. Their

school year was over, but Dorothy did get two applicants in response to a notice the school put on its bulletin board. One of them named Mathew, was in the top 10 percent of his class and had excellent recommendations from his teachers. The other, named Peter, had a mediocre academic record, and his letters of reference were only lukewarm. During the interview, however, Dorothy was more impressed with Peter than with Mathew. Peter examined the books quickly, made several useful observations about possible changes in the way they were kept, and spotted a mathematical error made by the previous bookkeeper. Mathew pored over the books for a long time and then asked several questions that struck Dorothy as somewhat naive.

Based on the information available to Dorothy, which applicant do you think she should hire? Why?

* Please consider the problem for a few moments. After you have considered the problem and analyzed it for a minute or two, turn the page for our analysis.

ANALYSIS

Dorothy is trying to draw a conclusion about two populations. One of the populations is composed of Mathew's bookkeeping ability and one population is composed of Peter's bookkeeping ability. Each population can be divided into two categories: one category being excellent bookkeeping abilities, and one category being less than competent bookkeeping abilities. Dorothy's conclusion that Peter was the better bookkeeper was based upon observing a sample of size 1; the 1 occasion Dorothy interviewed Peter and Mathew.

The Law of Large Numbers states that small samples are less reliable indicators of the population than are large samples. Dorothy had only seen a sample of size 1 tests of bookkeeping ability and this sample is so small that any conclusions drawn from it are very unreliable. Thus, Dorothy's claim is quite unwarranted: she would need a larger sample of tests to really know about the comparative abilities of the two applicants. The academic records, on the other hand, are a better indicator of the applicant's ability's: the applicant's academic record would have been based a larger sample of tests than Dorothy's sample of size 1, and the teachers who wrote the recommendations also would have had observed a larger sample of the applicant's bookkeeping abilities than Dorothy's 1 observation.

Re-paired**Concept 4 (large sample but unrelated pop) and Concept 3 story line**

Denise was the head of a recreation centre for children. This year she let the old bookkeeper go in order to hire someone who would manage the records using the Speed Organization System (SOS). There was a small accounting school in the area which Denise called. Their school year had ended, but Denise did get two applicants in response to an ad the school put on its bulletin board. One of them named Janet, was in the top 10 percent of her class, and had excellent letters of recommendation from her teachers stating the quality of her expertise with other organization systems. The other applicant named, Terri, had received average marks and her application had arrived late suggesting to Denise that Terri didn't really want the job. During the interview, however, Denise was more impressed with Terri than with Janet. Terri used SOS quickly and effectively, and even suggested how it could be used to correct errors made by the previous bookkeeper. Janet had trouble with SOS as she was slow, made mistakes, and asked several questions that struck Denise as somewhat naive.

Based on the information available to Denise which applicant do you think she should hire? Why?

Concept 3: Small sample pitted against a large sample

The president of a manufacturing company found that sales were down 60% this year. This decrease in revenue made it very difficult for the company to continue employing the number of people currently on the payroll. Cuts had to be made. Since sales in North America had dwindled down to essentially nil, Chris, the office manager, suggested that knowledge of the European market, as reflected in the employee's total number of sales abroad, would be essential when considering who to keep and who to let go. Johnson had worked at the company for eight years and was regarded by the managers to be a valued employee. He was efficient, cautious, and his knowledge of the European market was good as his sales were always well above the break even mark. The office manager's policy of gauging an employee's knowledge of the European market in terms of one's sales record would prove to be disastrous for Johnson because Smith, upon arriving at the company one year ago, had closed a very large number of sales deals recently. Smith was efficient, but unlike Johnson, was not very cautious in his dealings. The office manager decided to keep Smith and let Johnson go.

Do you think Smith should have been retained? Why? Explain your answer.

* Please consider this problem for a few minutes. After you have considered the problem and analyzed it for a minute or two, turn the page for our analysis.

ANALYSIS

The office manager is trying to draw a conclusion about two populations. One of the populations is composed of Smith's sales ability and one population is composed of Johnson's sales ability. Each population can be divided into two categories: one category being above average sales ability, and one category being below average sales ability. The office manager's conclusion that Smith had a better sales record was based upon observing a sample of size 1; 1 year of sales.

The Law of Large Numbers states that small samples are less reliable indicators of the population than are large samples. The office manager had only seen a sample of size 1-- this sample is so small that any conclusions drawn from it are very unreliable. Thus, the office manager's claim is quite unwarranted: he would need a larger sample to really know about the comparative abilities of the two employees. Using the Law of Large numbers, one could be more certain about Johnson's knowledge of

the market as his sales record was based on a sample of size 8 years worth of sales.

Re-paired

Concept 4 (large sample but unrelated pop) and Concept 3 story line

The annual report informed the president of a clothing retail company that sales had plummeted 65% this year. In order to get through this crisis, the company would have to downsize its operations in Canada, and it was decided that people who were not making very many sales in Europe would be let go. The company's accountants noted that clothing sales in Canada had decreased because of the recession, and since the economic forecast called for more of the same, the European market would be the company's major source of revenue in the coming years. Jackson a valued employee for the last ten years was efficient, cautious in his business dealings, and the revenues obtained in his domestic sales were always well above the break even mark. Indeed, the president of the company was very fond of Jackson as he was responsible for the ideas behind the organization of the company's infrastructure. Schmitt, who arrived at the company one year ago, obtained a very large number of sales in Germany recently, but unlike Jackson he was not very efficient. The president of the company did not value Schmitt as he did Johnson, but with great regret let Jackson go and kept Schmitt.

Comment on the president's decision.

Concept 4 : Large sample taken from a different population

Sandy was a senior at a music academy who was about to take the National Music Aptitude Test. It was important for her to do well on the NMAT because she wanted to get into a very selective music conservatory on the East Coast, one that required that applicants score at least 600. One night, Sandy sat down in her bedroom--a very relaxing environment--and completed a practice NMAT consisting of 200 questions. She scored the equivalent of 650. The next morning at school, one of her music teachers gave out a 15-question practice test from the NMAT in a setting that was designed to simulate the conditions she would face when she took the real test. On this, she scored a 550.

What do you think will happen when Sandy takes the real NMAT? Do you think she will score at least 600, or not? Explain your reasoning.

ANALYSIS

Sandy wants to draw a conclusion about a population. We can define the population as the all the questions on the real NMAT test. By answering a sample of NMAT questions, Sandy hopes to be able to predict how well she will do on the real NMAT on the basis of how well she did on the practice tests. The Law of Large Numbers states that larger samples are more reliable indicators of the population than are smaller samples. From this rule, one would expect Sandy's score on the NMAT to be at least 600 since she scored the equivalent of 650 on a practice test consisting of 200 questions. This however is not the correct answer.

In order to draw conclusions about a population, the sample must be taken from the population that you want to make conclusions about. The teacher's 15 questions came from a NMAT test that simulated the conditions Sandy would face when she took the real test. Sandy's 600 questions came from a NMAT test that did not simulate the conditions of the real test. Therefore, even though the sample size of the questions on Sandy's practice test was larger than the sample size of the teacher's practice test, the results of her test cannot be used to predict how she will do on the real NMAT test.

Concept 4 : Large sample taken from a different population

The director of a production of Shakespeare's *As You Like It* had just finished interviewing candidates who would act as the stage manager for the show. The person for the job had to know the play by heart as it was their job to make sure that the staging captured the subtle nuances of the play. The director read parts of the play and asked the candidate to answer questions regarding atmosphere and interpretation. The director liked two of the candidates very much. They were able to answer all three questions about the play that she posed, and more important the detail they gave was exceptional. Another candidate for the job was a stage manager whom the director had worked with before in three other Shakespeare comedies. The director thought she had done a superb job in each. Unfortunately, of the three questions posed to her about this play, one had been answered fairly well, but two lacked detailed information. This third candidate had to know immediately whether she was going to be chosen for the job. If not she would take another job in a play that would keep her busy for the next six months.

What should the director do - hire the third candidate or one of the two whose interview performances he liked better? Why?

* Please consider this problem for a few minutes. After you have considered the problem and analyzed it for a minute or two, turn the page for our analysis.

ANALYSIS

The director is trying to draw a conclusion about a population. The population is defined as the candidates' knowledge of the play *As You Like It*. According to the Law of Large Numbers, the larger a sample is the more reliable an indicator of the population it is. The director had observed the third candidate in three other Shakespeare comedies, but had only observed the first two candidates' answers to the three questions posed during the interview. One might think that the third candidate should be chosen since the director had observed a large sample of this candidate's work and her impression of this work was very favorable. This is not the correct answer. The director should choose one of the two candidates whose interview performance she liked better.

In order to draw conclusions about a population, the sample must be taken from the population that you want to make statements about. The director wanted to specifically assess the candidates' knowledge of the play *As You Like It*, and in order to do this she sampled their knowledge of this play by asking each candidate three questions about the play. The third candidate clearly did not demonstrate a good knowledge of *As You Like It* as reflected in her answers. The third candidate's knowledge of other Shakespeare comedies is irrelevant because the director is interested in the

candidate's knowledge of this particular play. Therefore, the differences in sample size do not matter in this situation.

The following problems were only seen at training

Concept5 : Regression towards the mean

John was a math professor who had been teaching the same introductory calculus class for close to thirty years. He talked to a colleague about a curious phenomenon he had noticed over the years.

"I give a 15-minute quiz every week. I make them very difficult--the average is rarely above 65%--because I believe in challenging the students. In addition, I want to identify the brilliant students early in the term so that I can encourage them to become math majors. Typically, after the first two weeks, the highest average in the class is about 95%. But in the thirty years I've been teaching, only about two or three have been able to maintain that average over the entire course. The rest of the highest students have ended up with lower averages."

Why do you think this is? Why is it that the student with the highest average after two weeks is almost always unable to maintain a 95% average over the entire terms? Explain your answer.

* Please consider this problem for a few minutes. After you have considered the problem and analyzed it for a minute or two, turn the page for our analysis.

ANALYSIS

We can analyze this problem using the Law of Large numbers in the following way. Each quiz can be seen as a sample of a student's knowledge about calculus. But what is the population? Roughly speaking, the population is the student's "true" level of understanding about the material covered in the quiz, or "true" level ability at calculus. For some students, their true ability is very high; for others it is very low. Because the quizzes are very difficult, the average ability seems to correspond to 65%. It is unlikely that there are students whose true ability is higher than 90%, and probably no one whose true ability is as high as 95%.

During the first two weeks, students have taken only two fifteen minute quizzes. This is quite a small sample. We know from the Law of Large Numbers that small samples are unreliable estimates of the population. It is entirely possible with a small sample that a student can score considerably higher than his or her true ability. But at the end of the term, the student's average consists of a much larger sample--fifteen or more quizzes. We should expect the student's average at the end of the term to be a more reliable indicator of the student's true ability for calculus. Those students who happen to score 95% after the first two weeks cannot maintain

that average over the entire term. Thus, it is not surprising that after two weeks, there may be some students whose average is 95%, but that very few students have an average that high at the end of the term.

Concept 6. Theory not founded on data

Researchers at an educational research institute recently conducted a study that examined the effect of caffeine on SAT scores. They recruited 500 high school students. Half of the students were given three cups of caffeinated coffee just before taking the SAT; the other half was given three cups of decaffeinated coffee. From the results of the study, the researchers concluded that caffeine had no detectable effect on SAT scores: the scores of students given caffeine did not differ at all, on the average, from the scores of students not given caffeine.

Just after the study was completed, an educational writer for a New York newspaper criticized the study. He stated that it was obvious that caffeine has a negative effect on a student's performance. According to him, the arousal created by the caffeine would make it difficult for students to concentrate on the questions, thereby leading to lower scores. The writer stated that the conclusions of the study had to be wrong.

What do you think of the newspaper writer's criticism of the study? Do you think it is well-founded? Explain your answer.

* Please consider this problem for a few minutes. After you have considered the problem and analyzed it for a minute or two, turn the page for our analysis.

ANALYSIS

The goal of the study was to determine whether caffeine has a detrimental effect on SAT scores. The researchers took a sample of 500 students in order to test this hypothesis. This sample was drawn from the population of high school students. On the basis of the sample, the researchers made an inference about the population: that there was no effect of caffeine on SAT scores.

The Law of Large Numbers states that as the sample size increases, the sample becomes a more reliable indicator of the population. A sample of 500 is quite a large sample. The inferences that are made from a sample this large are likely to be very reliable. Apart from other considerations, we should trust the findings of the study.

What about the educational writer's theory? Although his theory is plausible, the opposite theory is equally plausible. One might expect that caffeine would have a positive effect on SAT scores since a little bit of arousal may serve to increase concentration. The point here is that a

plausible theory that is unsupported by data is not a good basis for judgments, especially when there is a large sample that contradicts the theory. Thus, the writer's reasoning is not sound: It is more likely the case that caffeine has no discernable effect on performance on the SAT.

We are interested in studying how people go about explaining and predicting events under conditions of very limited information. It is important to us to study how people explain and predict under these conditions because such situations occur very frequently in the real world. Indeed, we often have to make important decisions and predictions under conditions of limited information, either because there is too little time to get additional information or because it is simply unavailable.

Experts who study human inference have found that principles of probability are helpful in explaining and predicting a great many events, especially under conditions of limited information. One such principle of probability that has been found to be particularly helpful is called the Law of Large Numbers. In simple terms, this principle says that the more different pieces of evidence you have, the better picture you get of the true situation. It is easy to see the application of the principle in the domain of sports. For example, when assessing an athlete's ability, the more games you see him or her play (the larger the sample), the better the idea you get of the athlete's true ability. Large samples are definitely better than small ones. That is why the principle is called the Law of Large Numbers. You can apply this principle by analogy to other cases in which your evidence is like a small sample (where you cannot be very sure of the conclusions) or cases in which your evidence is like a large sample (where you can be quite sure of the conclusions).

To illustrate how the Law of Large Numbers can be used to understand events in the domain of sports, we next present six situations in which the

Law of Large Numbers applies. Each situation is then analyzed in terms of the Law of Large Numbers.

We are interested in studying how people go about explaining and predicting events under conditions of very limited information. It is important to us to study how people explain and predict under these conditions because such situations occur very frequently in the real world. Indeed, we often have to make important decisions and predictions under conditions of limited information, either because there is too little time to get additional information or because it is simply unavailable.

Experts who study human inference have found that principles of probability are helpful in explaining and predicting a great many events, especially under conditions of limited information. One such principle of probability that has been found to be particularly helpful is called the Law of Large Numbers. In simple terms, this principle says that the more different pieces of evidence you have, the better picture you get of the true situation. It is easy to see the application of the principle in the domain of ability testing. For example, when trying to determine whether a person is skillful on a certain task, the more information you have about that person's performance on that task, the better idea you get of that person's true ability. Large samples are definitely better than small ones. That is why the principle is called the Law of Large Numbers. You can apply this principle by analogy to other cases in which your evidence is like a small sample (where you cannot be very sure of the conclusions) or cases in which your evidence is like a large sample (where you can be quite sure of the conclusions).

To illustrate how the Law of Large Numbers can be used to understand events in the domain of ability testing, we next present six situations in

which the Law of Large Numbers applies. Each situation is then analyzed in terms of the Law of Large Numbers.