INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

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

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

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

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

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

Order Number 1342275

The effects of training on statistical reasoning

Jones, Scott Fariss, M.A.

Rice University, 1990

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

Rice University

The Effects of Training on Statistical
Reasoning

by

Scott F. Jones

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

Approved Thesis Committee

Nancy J. Cooke,
Assistant Professor of
Psychology, Director

William C. Howell,
Professor of Psychology

Randi C. Martin,
Associate Professor of
Psychology

Houston, Texas

April, 1990

The Effects of Training on Statistical Reasoning

by

Scott F. Jones

Abstract

The methods which people use to reason about everyday events and the strategies they employ have received much attention throughout the years. One aspect of this history is the debate about whether learning rules or examples most facilitates transfer of knowledge to a different domain. This research attempted to answer this question through two experiments. The first experiment concentrated on defining the dimensions along which subjects perceived problems which embodied statistical heuristics. The results identified a contextual dimension along which subjects classified the problems. The second experiment was conducted to determine if the contextual dimension or the problem domain dimension could best account for transfer of training to novel problems. The results indicated that the training transferred to all novel problems, however, training did not transfer to a different set of problems presented to the subjects as a phone survey. Explanations for this lack of transfer are discussed.

Acknowledgements

I would like to express my gratitude to those who made this research possible. First and foremost I want to thank Sue Parker who put up with my daily misgivings about this research and provided true bits of wisdom not only about my research but also about my life. My friends David Mayer, Kim Donner, Laura Yaffee, Shannon Halgren and the rest of the graduate psychology students deserve special mention for their constant encouragement and loyalty. And last, but certainly not least, to my parents, John and Anne Jones, who gave me the strength to start this project and had the courage and grace to let me see it through on my own. They are truly the best parents any child could hope to have.

Finally, I would like to thank those who guided this research. My advisor, Nancy Cooke was always available to talk about the research and gave me countless hours of her time and valuable comments in revising the many drafts of this thesis. My other committee members Bill Howell and Randi Martin were very helpful in the final reviews of the manuscript and in providing the thought provoking ideas for future research.

Table of Contents

Abstract	ii
Acknowledgements	iii
Table of Tables	v
Table of Figures	vii
Teaching Reasoning	1
Heuristics and Statistical Reasoning	2
Non-statistical Heuristics	2
Statistical Heuristics	3
Learning to Reason	7
Training and Statistical Heuristics	8
Experiment 1	14
Method	15
Results	22
Discussion	33
Experiment 2	35
Method	36
Results	43
Discussion	56
General Discussion	59
References	62
Appendix A	65
Appendix B	81
Appendix C	91
Appendix D	103
Appendix E	121
Appendix F	126

Table of Tables

Table 1	Novice Demographics - Descriptive Statistics	17
Table 2	Stimulus Organization and References	19
Table 3	Reading Grade Level Analyses of Stimulus Materials for the First Experiment	20
Table 4	Stress, Average Stress, Proportion of Variance Explained, and Constraint for Each of the Three Groups of Subjects by Individual Differences Scaling Solutions of Dimension 2 to 4.....	24
Table 5	Reading Level Analyses of Stimulus Materials for the Second Experiment	41
Table 6	Item Analyses for the 12 Test Questions Used as Stimuli in Experiment 2.....	45
Table 7	Training Conditions and Mean Differences for Test Problems Which Were the Same as the Training Examples and Those Which Differed from the Training Examples by One and Two Dimensions.....	50
Table 8	Mean Differences in Statistical Rating Averaged Across Same versus Different Test Problem Dimensions for Academic Objective (AOT), Sports Probabilistic (SPT), and Combined (CT) Training.....	51
Table D-1	Stimulus Organization and References for Experiment 2.....	104
Table F-1	Demographics - Descriptive Statistics	127
Table F-2	Test Problem Analysis: Analysis of Variance.....	128
Table F-3	Analysis of the Simple Effects for the Test Problem by Training Condition Interaction.....	129
Table F-4	Analysis of the Simple Effects for the Test of Transfer of Training.....	130
Table F-5	Test Problem Analysis: Sample Size, Means, and Standard Deviations for Test 1	131
Table F-6	Test Problem Analysis: Sample Size, Means, and Standard Deviations for Test 2	132
Table F-7	Phone Survey: Analysis of Variance	133
Table F-8	Analysis of the Simple Effects for the Test Phase by Training Condition Interaction.....	134

Table F-9	Analysis of the Simple Effects for the Test Phase by Test Problem Interaction.....	135
Table F-10	Phone Survey Analysis: Sample Size, Means, and Standard Deviations for Test 1	136
Table F-11	Phone Survey Analysis: Sample Size, Means, and Standard Deviations for the Telephone Survey	137
Table F-12	Phone Survey Analysis: Sample Size, Means, and Standards Deviations for the Test 2.....	138

Table of Figures

Figure 1	Expert MDS Analysis.....	28
Figure 2	Accomplished MDS Analysis Dimension 1 versus Dimension 2.....	29
Figure 3	Accomplished MDS Analysis Dimension 1 versus Dimension 3.....	30
Figure 4	Accomplished MDS Analysis Dimension 2 versus Dimension 3.....	31
Figure 5	Naive MDS Analysis	32
Figure 6	Main Effect of Training Condition for the Test Problem Analysis	46
Figure 7	Main Effect of Test Phase for the Test Problem Analysis.....	47
Figure 8	Main Effect of Test Problem for the Test Problem Analysis.....	48
Figure 9	Training Condition by Test Problem Interaction for the Test Problem Analysis.....	49
Figure 10	Main Effect of Test Phase for the Phone Survey Analysis.....	52
Figure 11	Test Phase by Training Condition Interaction for the Phone Survey.....	53
Figure 12	Test Phase by Test Problem Interaction for the Phone Survey..	54
Figure 13	Main Effect of Training Condition for the Phone Survey.....	55
Figure 14	Main Effect of Test Problem for the Phone Survey Analysis.....	55
Figure B-1	Expert Cluster Analysis	85
Figure B-2	Accomplished Cluster Analysis.....	86
Figure B-3	Naive Cluster Analysis.....	87
Figure B-4	Expert Network Analysis (Pathfinder Algorithm)	88
Figure B-5	Accomplished Network Analysis (Pathfinder Algorithm).....	89
Figure B-6	Naive Network Analysis (Pathfinder Algorithm).....	90

Teaching Reasoning

The methods which people use to reason about everyday events have received much attention, throughout the years. In fact, discussion about this complex cognitive ability can be traced to the works of Plato and Aristotle. Recently, psychologists have focused their attention on two types of reasoning: inductive reasoning and deductive reasoning. Inductive reasoning concerns the process of movement from specific ideas or instances to more general concepts, whereas deductive reasoning is primarily concerned with analyzing some set of data to arrive at a conclusion. Although deductive reasoning is more the domain of the logician and inductive reasoning more the realm of the statistician, both of these areas have their roots in the works of Plato and the philosophy of education.

For many centuries the fundamental assumption which educators made about teaching was that training in abstract systems of rules could prepare the mind for the solution of concrete problems. This statement was authored by Plato and has come to be known as the doctrine of formal discipline. It suggests that individuals trained in logic or mathematics will be more able to solve problems in the everyday world. However, this view came into disrepute during the early part of the 20th century with Thorndike's (1906) work on transfer of training between tasks. Thorndike found that the primary factor which determined transfer of training was the degree of commonality between two tasks. Further research in general problem solving has tended to support this view, particularly the work of Simon and Hayes (1976) and Hayes & Simon (1977) who found that training on the Towers of Hanoi problem does not generalize to other isomorphic problems.

More recent research in this area has led to a reconsideration of the importance of abstract rule systems. This revival began with Piaget's (Piaget & Inhelder, 1975/1951) work on child development. Piaget speculated that reasoning is managed by operations in deductive logic and by processes which allow the application of the formal rules to everyday situations. He also believed that these rules were formed very early in adolescence and were inferential in nature and could not be taught. Neither study of abstract systems of rules nor study of examples within a

specific domain would help individuals to acquire the abstract form of the rules: this could only happen through induction of the rules by active discovery. Although this latter idea has been questioned, many of the concepts which Piaget set forth are useful in understanding the processes which people use to reason about everyday events.

In this paper, I examine some recent work that suggests that formal training affects statistical reasoning. I do this in several stages. First, non-statistical and statistical heuristics are discussed. Second, the factors which are thought to affect the use of statistical heuristics are considered. Third, the possible cognitive structures that are associated with learning statistical reasoning are explored. Fourth, two experiments are described which address these concerns. I begin with a brief discussion of statistical heuristics.

Heuristics and Statistical Reasoning

Non-statistical Heuristics

Recent research by Kahneman & Tversky (1972) has led to the discovery of heuristics that people use to reason about probabilistic events. These are diverse rules of thumb which allow individuals to simplify their analysis of certain situations. Essentially, they are processes that satisfice. That is, instead of working to the optimal solution, an individual works to meet some satisfactory criterion; the goal of maximizing is replaced with the goal of satisficing. Two major heuristics which have been identified include the representativeness heuristic (Kahneman & Tversky, 1973, 1974) and the availability heuristic (Tversky & Kahneman, 1973).

The representativeness heuristic concerns the judgements people make when comparing a case in point with a concept. Individuals assess how close the case is to some prototype—a model which illustrates the typical qualities of a class—and make a decision based on that information. Some of the problems associated with the use of this heuristic are insensitivity to prior probabilities and sample sizes. The primary function of the availability heuristic is to assess the frequency of occurrence of an event. It is believed that the process is based on

the ease with which an object can be brought to mind. This heuristic is sensitive to the salience of the object being retrieved and the effectiveness of the search set (Tversky & Kahneman, 1973). Therefore it results in systematic errors such as stereotyping and overestimating the frequency of very salient events, such as murder, rape, and cancer.

Although these heuristics are commonly used by individuals and often lead to correct estimates, they have major limitations. Frequently, important consequences such as sample size and prior probabilities are ignored, while unimportant considerations such as salience of the event and effectiveness of the search set are used without question. These problems suggest that humans have some fairly serious limitations on their abilities to reason. Indeed, not only do novices make these mistakes, but experts are known to be affected by the biases inherent in these heuristics. For example, it has been found that medical doctors rarely use the base rate information given to them in making diagnoses about breast cancer (Eddy, 1982). Overall it appears that people do not make use of some critical information when reasoning using heuristics. However, is reasoning equally biased when using statistical heuristics as it is when using non-statistical heuristics?

Statistical Heuristics

As mentioned earlier, heuristics are rules of thumb which people use to solve complex problems. According to Nisbett, Krantz, Jepson, & Kunda (1983), "statistical" heuristics are "intuitive, rule-of-thumb inferential procedures that resemble formal statistical procedures" (p. 345). The range of application of these processes may be widespread, even if quite selective in actual use. Thus, unlike the heuristics discussed by Kahneman and Tversky, these heuristics are based on statistical principles. Nisbett et al. (1983) have argued that people use these processes to reason about statistically uncertain situations.

Nisbett and his coauthors mention four basic types of statistical heuristics: the law of large numbers, regression to the mean, the base rate phenomena, and generalizing from instances. This paper will deal with two of these topics--the law of large numbers and regression to the mean.

Law of Large Numbers

The law of large numbers concerns the observation that larger samples from a population are more stable than smaller samples. For example, it is much easier to estimate the proportion of males born in a hospital with a large number of births, than in a hospital with a small number of births. This principle has been researched extensively by Fong, Krantz, & Nisbett (1986). They found that people could use the law of large numbers in reasoning about common, everyday events. Also, they found that the frequency and quality of statistical answers was influenced by training in statistics. For instance, the authors found that varying the nature of the training (either rule or example based) had a significant impact upon the kind of statistical answers given by the subjects. As Fong and his coauthors point out, "there appears to be a three-level ordering such that either rule or example training alone improves performance and that training on both has an additional effect" (p. 266). However, varying the type of training (subjective, objective, or probabilistic) had no effect upon the kind of statistical answers given.

Other research on this heuristic has been conducted by Nisbett et al. (1983). In a series of experiments they concluded that people were able to generalize from instances, and that subjects "exercise more caution when generalizing from single cases when heterogeneity is expected" (p. 348). In their first study subjects were given one instance of an object sampled from the population which had a certain property. Subjects were asked to evaluate the percentage of the population which would have that property. Different levels of heterogeneity were obtained by using the physical characteristics of metals, color of animals, and so on. They found that large sample size is "important primarily in the case of populations whom subjects believe to be heterogeneous with respect to the property in question" (p. 348).

One instance of the use of this statistical heuristic is provided by Nisbett and Ross (1980) who infer that "many lay concepts and maxims reflect an appreciation of statistical principles." (Nisbett et al., 1983, p. 342). They present several maxims which indicate a knowledge of the law of large numbers; for example "Don't judge a book by its cover" and "All that glitters is not gold."

These maxims indicate a basic underlying knowledge of the law of large numbers. Nisbett et al. (1983) provide another example of the application of statistical reasoning in everyday events. This example concerns a thought experiment. Specifically, one would not find it unusual to hear a grandmother say that she was surprised that all nine of her grand children were male. However, for a grandmother to be surprised that all three of her grand children were male is not as common. Presumably this is due to some form of statistical reasoning.

Regression

Another basic statistical heuristic concerns the principle of regression to the mean. If the value of one sample deviates from the mean of the group, then subsequent samples should deviate in the opposite direction. In other words when someone scores higher or lower than the group on some test, then on the next test that person will have a less extreme score and this score is said to have regressed toward the mean. Many common examples of this heuristic are found in sports and in testing. For instance, in most sports the rookie of the year traditionally has a worse second year or the team that wins the World Series may lose the first game of the next season. Also, in testing, when one selects a sample on the basis of high test scores, then, on subsequent administrations of a similar test, scores should decrease. Both of these situations illustrate the regression effect. Nisbett and Ross (1980) seem to feel that "beginners luck" may indicate some primitive understanding of this regression effect.

Further support for the use of the regression principle is found in a study from Lehman, Lempert, and Nisbett (1988). They conducted both cross-sectional and longitudinal studies of graduate students abilities to reason using statistical-methodological techniques including the regression principle. The graduate students were chosen from four disciplines: medicine, psychology, chemistry, and law, and were tested on "everyday-life" and "scientific" problems from each of the three domains. The authors found that statistical-methodological reasoning increased dramatically during graduate education (nearly 70%) for psychology students. From

this it can be inferred that people who are trained in statistical reasoning can use the regression principle in reasoning about everyday problems.

Factors which Influence the Use of Statistical Heuristics

Thus, statistical heuristics are used by people to reason, although they are not applied universally. What specific factors affect this selective application of the statistical heuristics? Is it the concrete nature of the situation? What about the cultural milieu in which a child is reared? Are there any effects due to the description of the sampling process? Is training an effective method for imparting this type of knowledge? These questions address different aspects of the problem and all are pertinent to the issue. Though the cultural factors are interesting in themselves, any experimental test of these factors would require access to a diverse population of cultures. Thus, I chose to focus on the effect of training on statistical biases and to temporarily set aside the cultural issues. This should not pose a problem if subject selection is carefully controlled to provide cultural homogeneity.

One factor which influences the use of statistical heuristics is the clarity of the sample space—do people understand what knowledge is necessary to solve the problem? One of the greatest difficulties in research on heuristics is that the sample spaces are unclear and the difficulty of repetition is great. This type of difficulty is manifested in the idea that brief exposure to an individual constitutes a complete portrait of the individual, rather than a sample of attitudes and abilities (Nisbett et al., 1983). Another factor which has influence on statistical reasoning is the degree of randomness of the problem and the cues which signal randomness. Training is also an important factor in the use of statistical heuristics and is of direct interest to this research. As the work of Fong et al. (1986) shows, training does influence statistical reasoning. However the exact manner of influence is unknown.

Although Nisbett (Nisbett, Fong, Lehman, & Cheng, 1987) suggests that abstract rules are learned. However, it is possible that concrete examples of problems are learned, and that analogical transfer is used to apply the solutions associated with these learned problems to novel

ones. If abstract rules are learned, then this knowledge should transfer to all problems that involve the particular rules. However, if examples are learned, then the knowledge should transfer only to the extent that the new problem shares features with the example, at least in the early stages of learning when only a few examples are known. Hence, the better the match the more likely the transfer to the novel problem. These possibilities are explored in depth in the next section.

Learning to Reason

As I mentioned at the beginning of this paper there are two major views about how people learn to reason. The first of these, espoused by Plato and Piaget and now Nisbett, implies that abstract rules are critical to learning. The exact method of rule acquisition is debatable, but the primary tenet remains: abstract rules are responsible for peoples' reasoning. The second view, supported by Thorndike and Simon, among others, argues that learning is a domain dependent function, predicated on knowledge of one content area. These two radically opposed philosophies have provided the basis for much heated discussion, one might even say argument, over the last 2000 years.

Some of the early research in cognitive psychology has focused on strategies used to solve a variety of problems. The seminal work of Newell & Simon (1972) has provided an original framework for understanding these strategies. They propose that a problem solver represents the problem under study by a space, called the problem space, which contains operators for moving from one state in the problem to another. As Howell and Cooke (1989) summarize, some of the general strategies which are used to search through the problem space include subgoal analysis and working backwards. Verbal protocols from subjects substantiated the use of these general strategies. This view maintains that general rules are used to solve problems. Similarly, Nisbett and his coauthors (Nisbett et al., 1987; Nisbett et al., 1983; and Lehman, Lempert, and Nisbett, 1988) think that abstract rules, in the form of statistical heuristics, guide successful statistical reasoning.

At the other end of the spectrum is research which emphasizes domain specific knowledge in problem solving. As Howell and Cooke point out, in their 1988 review of training in cognitive psychology, current research suggests that “problem solving skill is tied to domain-specific knowledge and the organization of that knowledge in memory” (p. 25). For instance, Chase and Simon (1973) found that certain groups of chess pieces, those forming attack or defense positions, were grouped together at encoding for chess masters but not for novices. Also Egan and Schwartz (1979) found that skilled electronics technicians recall the elements of a circuit diagram in functional chunks. Further, Adelson’s (1981) study of programming languages revealed different subjective organizations of programs in novice and expert programmers. These studies suggest that an expert’s skilled problem solving is linked to a body of domain specific knowledge which novices simply do not possess. In summary, learning to reason can be described as acquiring abstract rules and applying them to a variety of problems or learning concrete examples and extending these instances to novel problems.

Thus, given that statistical reasoning can be learned as Nisbett suggests, then investigating the aspects of training that are critical may shed light on how this statistical knowledge is represented—abstract rules, concrete instances or both. If people learn abstract rules then direct training on abstract rules should result in better performance in comparison to training on specific examples. Similarly, if people learn instances then training on examples should result in better performance. Fong et al. (1986) carried out a study to address the specific aspects of training that benefit performance.

Training and Statistical Heuristics

In their first experiment Fong et al. (1986) investigated the effects of type of training on statistical reasoning related to the law of large numbers, as well as the effect of type of training on their responses. The subjects were asked to read 18 problems, think about the problems, and then write down reasonable answers to the problems. For each problem a common, everyday situation was introduced to the reader and the protagonist of the story made a decision about the

situation. That decision either expressed good or bad judgement depending on whether or not the correct statistical principles were applied by the protagonist. The last sentence of each problem then asked the subject to comment on the analysis made by the protagonist of the story. The answers given by the subjects were recorded and then later analyzed for both frequency and quality of statistical answers.

The themes for each of the 18 problems used in the study were constructed by crossing problem type with problem structure. The type of problem was composed of three levels: probabilistic, objective, and subjective (which the authors refer to as different domains.) The cover stories included purchasing new cars, sports, academics, and testing. Probabilistic problems were generated by incorporating some type of random variation into the body of the story. This was done by stating the range of the sample outcomes, by including a random generating device, or by stating that the sample was random. The objective problems were created by including neutral cover stories about random events. For these problems no randomness cue was stated. Subjective problems stressed a non-random event which was embodied in the story. For example, one of these non-random events described a student who had to make a decision about college based either on one subjective experience or his friends' year long exposure to the college.

In order to provide some similar structure and more variance, the three types of problems were crossed with six types of structure. Essentially, Fong and his coauthors invented a taxonomy for distinguishing between problems used in statistical heuristics. The 6 structures they developed pitted different varieties of populations and samples against each other. Subjects were then required to make inferences based on the differences between the samples and populations represented by each structure. For instance, structure 1 required subjects to draw conclusions about a population from a single small sample, whereas structure 2 pitted a small sample against a large sample. Problem C in Appendix A is an example of a problem written with structure 1, while problem H in Appendix A is written with structure 2. Structures 3 and 6 were

both false alarm problems which were designed to see if subjects would selectively apply the law of large numbers heuristic or if they would apply it to every problem they read. These 18 problems were thought to provide a stable basis from which comparisons of subjects' answers could be contrasted.

Subjects were divided into five groups, each group receiving a different type of training on the law of large numbers. The five types of training received by the subjects were rule-based, example-based, full—both rule-based and example-based, demand, and control. In the rule-based group, the theoretical foundation of the law of large numbers was reviewed. The group which received example-based training was given three examples of problems which incorporated the law of large numbers. In the group which received the full training, the subjects were first given the rule-based training, followed by the example-based training. Subjects in the demand condition were given a single sentence definition of the law of large numbers and a brief example. This condition was included as a check on the demand characteristics of the study. That is, we were concerned with whether subjects gave statistical answers because statistics had been discussed in the first part of the study. The subjects in the control condition received no training before they answered the 18 questions.

Frequency and quality of statistical reasoning differed significantly for training types and problem types, while the training by problem interaction was only marginally significant. In general, training was better than no training (control or demand), and full training produced more statistical answers than example or rule training. The frequency of statistical answers was not reliably different for both the Control and Demand groups or for both the Rule and Example groups. The frequency with which the Full training group produced statistical answers was significantly greater than any of the no-training or single training groups. The quality of statistical answers was similar for the Control and Demand groups. However, the quality of the statistical answers did not differ for any of the training groups—Rule, Example, or Full. These results do not help to shed light

on the rule versus example controversy because lack of control of training time makes the results difficult to interpret.

Fong et al's (1986) second experiment was designed to investigate transfer of training on problems from a specific domain (referred to as problem type in experiment 1) to problems in the same or a different domain. Subjects were assigned to one of four conditions. In the first three conditions subjects were given training similar to that in the full training condition of experiment 1, but the type of the example problems was varied. In the probabilistic condition subjects received three examples of probabilistic problems. Similarly, in the objective and subjective conditions, the type of examples were, respectively, objective and subjective. The fourth condition, the control, was identical to that of experiment 1. After subjects were trained they received the same set of 18 problems as in experiment 1 and were tested with the same procedure. If specific examples are learned, then training should only transfer to problems similar to the training problems.

Fong et al. (1986) found that the quality and frequency of statistical answers in a testing domain was not affected by training domain. That is, training on probabilistic problems transferred to probabilistic, objective, and subjective problems. Thus Fong et al. (1986) conclude that "training significantly increased statistical reasoning; the domain of training had no differential effect" (p.273). These results support the idea that general rules are learned, however several aspects of the methodology make this conclusion questionable.

Although, Fong et al. (1986) concluded that training transfers to different domains, the domains were defined by Fong! If subjects do not perceive these differences then the conclusion that training is "domain independent" is questionable. In order to resolve this dilemma, the problem features which are salient to the subjects should be determined. If the subjects do perceive the features which Fong et al. (1986) used then the domain independence of training is supported. However, if they use another categorization scheme, then the question of transfer of training should be analyzed in light of the new features.

Another problem with this study (Fong et al., 1986) and others (Fisk & Gallini, 1989; Agnoli & Krantz, 1989; Nisbett et al., 1983, 1987; Lehman et al., 1988) is that of demand characteristics. Simply bringing subjects into a laboratory setting and asking them to "reason" may influence their behavior. Although training in the laboratory may transfer to tests taken in the laboratory it may not generalize to everyday reasoning outside the laboratory. This research seems especially susceptible to demand characteristics because there are no specific behaviors which can be observed from subjects' actions when they are "statistically reasoning." In order to rule out demand characteristics as an explanation for subjects' responses, there must be some explicit test of reasoning ability outside of the laboratory. One method of circumnavigating this problem is to ask subjects to answer statistical questions in the guise of an attitude survey.

Fong et al. (1986) used this approach in a separate experiment. They randomly selected 193 male students from the introductory statistics course at the University of Michigan. Half of the subjects were contacted during the first week of classes and the other half were contacted during the last week of classes. The subjects were told that they were participating in a survey of students' opinions about sports. After answering several questions about their own demographics and a few sports controversies, the subjects were asked several questions which had an underlying statistical basis, for example the Rookie of the Year problem. They found that training had a significant impact upon the use of statistical explanations for the questions. Although this approach demonstrates the validity of the opinion survey methodology, it does not assess the relative degree of transfer of training. Another major concern is that no females were tested in the experiment.

Another method which has been used to rule out the effect of demand characteristics is the use of a 'demand' or control condition (Fong et al., 1986). Subjects in this condition were given a "one sentence definition of the law of large numbers that introduced the examples training, along with a brief example. " (p. 261). This condition may allow for the control of experimenter demand, as Fong et al., (1986) report, but it does not control for the effect of the

laboratory setting. Thus, they do not address whether training transfers in the general context, but only whether it does so in the laboratory setting.

In essence, there are two questions which I sought to answer through two experiments. The first question concerns whether formal statistical training transfers to different problems (supporting rule learning) when the differences are subjectively defined. Research conducted in the Kahneman and Tversky tradition indicates that even statistical experts are subject to bias in statistical reasoning suggesting that their training does not transfer. However, some of the work done by Fong seems to indicate that training does have an effect on the solution of problems in different domains. How can these conflicting findings be resolved and what light do they shed on what is learned during training? Perhaps the definition of a "different problem" is at the heart of this conflict. If abstract rules are learned then training should transfer to all problems that the subject perceives as novel. If, on the other hand, concrete examples are learned, then transfer of training should occur to the extent that features of these examples map onto features of previously learned concrete examples.

The second question concerns whether training transfers to different problems (or even the same type of problem) when the potential for demand characteristics is reduced. Simply put, if demand characteristics play an important role in influencing statistical reasoning, then much of the past research in this area will have to be re-examined in a more critical light. It may be that subjects' reasoning abilities in the laboratory have no analog in everyday situations. If this is the case, the extent to which reasoning in the laboratory differs from that in the common world will have to be thoroughly examined before any general statements about transfer of statistical training can be made.

In summary, I took Fong et al's (1986) second approach to addressing the rule versus example question. In doing so there were two main hypotheses. First, in the laboratory does training transfer to different problems (supporting rule learning), when the differences are defined

by subjects? And second, does training transfer to different problems (or even the same type of problem) when the potential for demand characteristics is reduced?

Experiment 1

The primary purpose of this study is to investigate subjects' structure of the relationships of statistical problems such as those used by Fong et al. (1986) in their first experiment.

Essentially, the authors imply that there are three types of problem domains and six problem structures which characterize questions about statistical heuristics. More specifically, domain independence is defined in terms of the three problem domains so that transfer to novel problems is said to occur if training in one domain facilitates problem solving in another. However, it is not clear that problems in these different domains are actually perceived differently by subjects. Therefore, in order to define the problem space more clearly, Experiment 1 determined how individuals classify statistical problems.

In order to classify the problems in statistical heuristics the subjects must have a set of problems which is as complete as possible. To maximize the completeness of the set of problems presented to the subjects, a comprehensive search of the literature was carried out. This review included much of the literature on expert and novice differences (Chase & Simon, 1973; Chi, Feltovich, & Glaser, 1981; Cooke & Schvaneveldt, 1988; de Groot, 1965; Egan & Schwartz, 1979; Kaiser, Jonides, & Alexander, 1986; Newell & Simon, 1972; Reitman, 1976; Schoenfeld & Herrmann, 1982; Simon & Chase, 1973). In addition to the three domains used by Fong et al. (1986), two other dimensions, both suggested by Chi et al. (1981) in their investigation of physics problem solving, were chosen: 1) the type of statistical principle (e.g. the law of large numbers or regression) underlying problem knowledge; and 2) the problem context or surface features of the problem (in this case sports and academic). Problems were varied across these three dimensions and several additional problems were added which did not easily fit into any of the above areas, but which were relevant to the question. It was hoped that these three factors (domain, principle, and context) would define a rather complete space of problems in statistical heuristics.

In addition to investigating subjects' organization of this set of statistics problems, in Experiment 1 an investigation of expert and novice differences was also carried out. Specifically, novices should sort the problems according to surface structure (context) and experts should sort the problems according to deep structure (domain, structure, and principle). Both Schoenfeld & Herrmann (1982) and Chi et al. (1981) found that experts tend to sort problems by the underlying structure involved in the problem set. Schoenfeld & Herrmann (1982), in their study of mathematical expertise, found that experts sorted a set of mathematical problems according to the underlying principle embedded in the solution of the problem. The novices tended to sort the problems by syntactic similarity. In Chi et al.'s (1981) study of physics expertise, the experts sorted the physics problems according to the type of physical principle involved in the solution of the problem, while novices sorted the problems according to the surface context—wording or pictorial similarities in the representation of the problem. These results suggest a sensible hypothesis: novices should organize the problems using surface structure (i.e. context) and experts should organize the same problems using deep structure (i.e. principle). Support for this hypothesis would raise problems for Fong et al.'s (1986) interpretation of transfer to novel problems in which novelty was defined by the domain dimension.

Method

Subjects

Twenty-nine undergraduates from the Rice University Psychology Subject Pool volunteered to participate for course credit. The average age of the subjects was 19.8 years. The standard deviation was 2.9 years. The youngest subject was 18 years of age, while the eldest was 34 years of age. Fifteen males (52% of the subjects) and fourteen females (48%) participated in the experiment. The majority of students (62%) had not received any formal statistical training. Thirty-five percent of the subjects had taken one or two statistical courses. Thirty-eight percent of the subjects were majoring in the social sciences. Complete demographic statistics for the subject sample are reported in Table 1. By selecting subjects from the Rice subject pool, the

degree of cultural homogeneity should be maximized thus avoiding problems associated with different cultural prescriptions towards statistical reasoning.

Six Professors or graduate students in Psychology served as the experts. All of the expert subjects had completed several courses in statistics and 3 had extensive experience in applied statistical settings. The average age of the experts was 31.2 years (S.D. = 8.2). All of the experts subjects tested were white and 2 out of the 6 experts were female. Three of the experts had extensive teaching experience, including statistics, test and measurements, and experimental design.

Table 1

Novice Demographics - Descriptive Statistics

Variable	Number	Percentage	Variable	Number	Percentage
Gender			Age		
Female	14	48	18	7	24
Male	15	52	19	10	35
Race			20	6	21
White	20	69	21	5	17
Black	5	17	34	1	3
Other	4	14	Year In School		
Major			Freshman	8	29
Business	4	14	Sophomore	9	32
Engineering	9	31	Junior	6	21
Humanities	3	10	Senior	5	18
Natural Sciences	2	7	Piles Generated		
Social Sciences	11	38	3	4	14
Statistical Courses			4	12	42
None*	18	62	5	7	24
1 Course	6	21	6	2	7
2 Courses	4	14	7	2	7
5 Courses	1	3	8	1	3
			9	1	3

* 2 subjects included in this category reported attempting one statistics course

Stimuli

The stimuli consisted of 15 stories: some selected from previous research and some created by the author. These stories are presented in Appendix A. The specific dimensions corresponding to each problem are reported in Table 2. The stories were varied factorially along problem domain (probabilistic, objective, and subjective), statistical principle (law of large numbers and regression), and context (sports and academic). Three “extra” problems were included in the task set to increase task complexity. Although problem structure was not varied factorially in the stories, several levels of the factor were included in some of the stories. That is, there was no systematic effort to choose problems which reflected certain structures. Of the fifteen problems given to the subjects, 7 had sports contexts, 7 had academic contexts, and 1 was neither sports nor academic. Eight of the problems explored the statistical principle of the law of large numbers, while 7 problems recounted the regression principle. Objective, probabilistic, and subjective domains were each included in 5 of the problems.

One possible problem with these stories involves the notion that subjects could perceive and sort the stories based on the cognitive complexity of the story. For instance, subjects might encode the overall difficulty of interpreting and analyzing the story. To reduce the possibility that subjects would encode irrelevant aspects of the stories, reading grade level analyses were conducted (see Table 3). Results indicate that the stories were similar in structure, vocabulary, and syntax. The average number of words per sentence ranged from 11.9 to 23.0, with a mean of 16.9. Three reading grade level analysis procedures—Dale-Chall (1984), Forecast (Caylor, Sticht, Fox, & Ford, 1973), and Fry (1968)—were chosen. The average reading level for the Dale-Chall, Forecast, and Fry methods across the fifteen stories is, respectively, 9-10, 8.7, and 7.9. These numbers are in fairly close agreement. In addition, for the Dale-Chall, Forecast, and Fry procedures, the reading grade levels for all 15 stories are within ± 2 standard deviations of the mean of the respective reading grade level procedure.

Table 2

Stimulus Organization and References

ID	Context	Principle	Domain	Structure	Problem Reference
A	Sports	Law of Large Numbers	Objective	1	Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). <i>Page 356-The Championship Selection Problem</i>
B			Probabilistic	2	Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). <i>Appendix A, Problem 7, rewritten</i>
C			Subjective	1	Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). <i>Appendix A, Problem 13</i>
D		Regression	Objective	3	Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). <i>Page 354 -The Football Problem</i>
E			Probabilistic	3	Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). <i>Appendix A, Problem 3, rewritten</i>
F			Subjective	3	Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). <i>Appendix A, Problem 15, rewritten</i>
G		Regression	Objective	?	Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). <i>Page 359 - The Rookie of the Year</i>
H	?	Law of Large Numbers	Probabilistic	?	Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). <i>Appendix A, Problem 2, rewritten</i>
I	Academic	Law of Large Numbers	Objective	?	Author
J			Probabilistic	1	Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). <i>Appendix A, Problem 1</i>
K			Subjective	2	Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). <i>Appendix A, Problem 14</i>
L		Regression	Objective	3	Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). <i>Appendix A, Problem 9</i>
M			Probabilistic	?	Author
N			Subjective	?	Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). <i>Appendix A, Problem 15, rewritten</i>
O		Law of Large Numbers	Subjective	?	Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). <i>Page 356-The Class Selection Problem</i>

Table 3

Reading Level Analyses of Stimulus Materials for the First Experiment

Story Identifier	Mean Words per Sentences	Percentage of One Syllable Words	Reading Grade Levels		
			Dale-Chall	Forecast	Fry
A	23.0	68.7	11-12	9.1	9.0
B	18.6	67.3	9-10	9.3	9.0
C	15.2	75.9	5-6	7.9	6.0
D	17.4	70.5	9-10	8.8	9.0
E	16.3	75.4	9-10	8.0	7.0
F	14.4	70.5	7-8	8.8	8.0
G	18.6	72.0	7-8	8.5	7.0
H	13.8	78.0	5-6	7.6	5.0
I	18.8	71.7	9-10	8.6	10.0
J	15.1	69.8	7-8	8.9	7.0
K	20.7	67.2	9-10	9.3	10.0
L	17.8	73.7	7-8	8.3	7.0
M	15.4	64.3	9-10	9.8	9.0
N	17.0	67.6	9-10	9.3	9.0
O	11.9	75.9	5-6	7.9	6.0
Mean	16.9	71.23	9-10*	8.67	7.87
Standard Deviation	2.8	3.9		0.64	1.55
* Mode					

Procedure

Subjects were tested in small groups, ranging from 3 to 5 people in each group. Each group of subjects received a short introduction to the experiment, instructions for completing the experimental task, and read and signed the informed consent contract. The subjects were told the following:

We would like you to carefully read and and sort the following short stories into piles. In order to do this you should read each short story very carefully and possibly, begin to solve it. You should not try to completely solve the problem, but only to understand it. Then we would like you to sort the short stories into piles based on their overall similarity. You should use whatever features seem appropriate to you when sorting the stories. The size of your piles can be either large or small. While sorting the stories into piles you may find that one story belongs in more than one pile. You may include a story in more than one pile. To do this simply write the letter of the story on a blank card and put it in the other pile. After you have finished sorting the stories into piles, I would like you to name each pile. That is, write down what each pile represents.

Once again, carefully read the stories and begin solving them. Then sort the stories into as many piles as necessary. After you have finished sorting the stories into piles, name each pile.

Sort by whatever makes the most sense to you.

Piles can be any size.

Essentially, the subjects were instructed to use a modified Q-sort procedure (Rosenberg & Kim, 1975; Rosenberg, Nelson, & Vivikananthan, 1968). In a standard Q-sort subjects are instructed to sort the stories into logical piles and then resort the stories. Inter-rater reliabilities are calculated to ensure stability of the matrices. Allowing subjects to place stories in more than one pile allows for non-hierarchical organizations.

The subjects were allowed as much as time as they need to complete the task (most subjects required from 30 minutes to 50 minutes to complete the procedure). Upon completion of the task the subjects were asked to complete a biographical information blank which requested basic demographic information and specific information on statistical experience. This completed the testing procedure.

Analysis

The piles generated by each subject were then reduced to a co-occurrence matrix. Each of the stories served as an index along both dimensions of the matrix. If two stories were placed in the same pile by a subject then a "1" was placed in the cell corresponding to the indices obtained from the two stories. Otherwise a zero was placed in the cell. A co-occurrence matrix was generated for each subject and served as the basic unit for the correlational analyses. All other analyses were carried out on the summed co-occurrence matrices.

Results

Four basic analyses were conducted on the reduced data: correlation, multi-dimensional scaling, cluster analysis, and network analysis. Each analysis was also conducted separately for groups varying in statistical expertise. For the purposes of these analyses the experts were defined as those 6 graduate students and faculty who had extensive statistical experience. The 13 accomplished subjects, which refers to those students who had received some statistical course work, were undergraduates from the Rice University Subject Pool. The 16 naive subjects, also from the subject pool, reported they had no previous statistical experience. These three groups of subjects form the core groups for the study of expert/novice differences.

Correlation

First, correlational analyses were conducted on the data. Essentially each subjects co-occurrence matrix was correlated with every other subjects co-occurrence matrix. This resulted in a 35 by 35 matrix of correlation coefficients. The correlations were computed using Spearman's Rank-Order correlation coefficient. For this analysis only the naive subjects were used. The correlations for each naive subject with every other naive subject were then averaged ($r_{nn} = 0.25$, $N=120$) as were the correlations for each expert with every other expert ($r_{ee} = 0.56$, $N=15$). Then the correlations for each expert with each naive subject were averaged ($r_{ne} = 0.28$, $N=96$). The naive-naive correlation was then compared to the expert-expert correlation by use of Fisher's Z transformation ($Z = 1.29$, $p = NS$). Although this analysis revealed no statistical differences in the

expert and naive co-occurrence matrices, this was possibly due to the small number of experts tested ($n=6$). However, there was a large difference between the absolute values of the correlations of the two groups which suggests that the experts agreed more among themselves than the naive subjects did. Also every expert-expert correlation was significant at $r > 0.4393$, $p < 0.0001$. This indicates that there was a high degree of inter-rater reliability among the experts. In addition, the average naive-naive correlation and the average naive-expert correlation were not significantly different, indicating that there was no particular naive categorization for the stories.

Multi-Dimensional Scaling

In general, multi-dimensional scaling is a group of mathematical procedures which can elucidate the underlying structure of a set of data matrices. Usually, the data matrices are proximity or co-occurrence matrices. Basically, a mathematical algorithm determines the configuration of points in n -dimensional space which best fits the proximity data. This is usually done by an iterative search algorithm which finds the path of steepest descent to minimize the stress of the configuration. In this case the stress of the configuration is a measure of the “badness of fit”, Kruskal and Carroll (1969), where larger values of stress indicate a worse fit. While these algorithms suffer from the characteristic deficiencies of all steepest descent algorithms—local minima and incomplete convergence, with proper safeguards they are valid.

The first set of analyses conducted on the data was a non-metric multi-dimensional scaling procedure called INDSCAL (SAS, 1989). This approach, known as the Individual Differences Scaling model, was developed by Carroll and Chang (1970) and uses the inherent differences in the subjects matrices to determine a unique orientation of the axes (Kruskal & Wish, 1978). The authors point out that this allows for simplified interpretation of the n -dimensional space. Separate INDSCAL's were conducted for each of the three groups for 2, 3, and 4 dimensional solutions. The stress, average stress, average proportion of variance explained, and constraint for each INDSCAL analysis is reported in Table 4. It should be noted that average stress in Table 4 refers

to the root mean square average of the stresses for each subjects co-occurrence matrix, while stress refers to the badness of fit of the model.

Table 4

Stress, Average Stress, Proportion of Variance Explained, and Constraint for Each of the Three Groups of Subjects by Individual Differences Scaling Solutions of Dimension 2 to 4

Group	Dimensionality	Stress	Average		Constraint
			Stress	R-Squared	
Naive	4	0.083	0.067	0.957	0.434
Accomplished	4	0.961	0.074	0.942	0.427
Expert	4	0.034	0.028	0.995	0.473
Naive	3	0.115	0.109	0.929	0.491
Accomplished	3	0.114	0.099	0.933	0.501
Expert		0.035	0.040	0.993	0.560
Naive	2	0.120	0.130	0.926	0.579
Accomplished	2	0.143	0.140	0.910	0.569
Expert	2	0.006	0.072	0.978	0.637

Basic criteria for selecting solutions

There are several key indicators to be examined when choosing the dimensionality of the solution to be interpreted. These indicators are the constraint of the solution, the absolute level of stress of the solution, plots of stress versus dimensionality, and the ratio of entities to dimensions. The idea of the constraint of a solution was developed by Isaac and Poor (1974) and hinges upon deriving the expected stress of a configuration based on random data. The constraint is defined as the expected stress minus the observed stress. As Isaac and Poor explain, the larger the value of constraint, "the more the configuration differs from a random solution involving a given number of points." So the main objective is to maximize the constraint. In their paper Isaac and Poor provide tables which allow for calculation of constraint for various dimensionalities and number of points.

The last three indicators mentioned above—absolute level of stress, plots of stress versus dimensionality, and the ratio of entities to dimensions—are interrelated indicators which

can help to distinguish among the various dimensional solutions obtained. As Kruskal and Wish (1978) point out, as a rule of thumb configurations with stresses lower than 0.10 are usually acceptable. Plots of stress versus dimensionality generally will show an "elbow" at the correct dimensionality, if the error is sufficiently low. Finally the ratio of entities to dimensions can affect the interpretation of stress. As Kruskal and Wish describe, "As a rule of thumb, if $I [\text{entities}] > 4 \cdot R [\text{dimensions}]$ the interpretation of stress is not sensitive to I and R ."

How do these techniques apply to the data of this experiment? First, the number of entities, in this case stories, is 15. According to Kruskal and Wish, for the stress to be invariant under entities and dimensions, the number of dimensions should not exceed three. Thus, since the 4, 5, and 6 dimensional solutions violate this rule, the 2 and 3 dimensional solutions should be focused on. Second, the constraints for the 2 dimensional solutions for all three groups of experts are higher than their 3 dimensional counterparts (see Table 4), although there is no way to tell whether this difference is significant. However, the 2 dimensional solutions show rather high values of stress—ranging from 0.13 to 0.14—for all groups except the experts. The three dimensional solutions show somewhat lower values of stress, all below the 0.10 cutoff, with the exception of naive group. The stress for the naive group is 0.109, which is very close to the cutoff point and does not terribly violate the previous assumptions of invariance. Thus the 3 dimensional solutions, for all expertise sub-groups, meet the criteria of stability (or at least invariance with respect to entities and dimensions), high constraint, and low stress.

Interpretation

The interpretation of multi-dimensional scaling diagrams is at difficult. Since the 3 dimensional solution was chosen for all groups there will be three plots per solution. I have chosen to discuss the interpretation of the INDSCAL model for each group separately and then later contrast and compare the three groups. The three dimensional solution for the experts is particularly good. The configuration explains 99.3% of the variance in the subjects' data and the

stress was quite low, approximately 0.04. This indicates there was an excellent fit between the model and the data.

Figure 1 is a plot of expert dimension 1 and 2. As can be seen, dimension 1 clearly distinguishes between the statistical principles of regression and the law of large numbers. Positive stimulus coordinates indicate the principle of the law of large numbers and negative coordinates indicate the regression principle. This dimension also had a very high importance rating, 0.78, indicating that subjects heavily relied on this dimension. This is expected since experts should be able to see the underlying statistical principles clearly. The meaning of the second dimension is less obvious. It may represent the type of context of the story. Negative stimulus coordinates indicate primarily academic problems, while positive stimulus coordinates indicate mainly sports problems. In spite of the fact that several stories are mis-classified, it seems reasonable to interpret this dimensions as context. It should be noted that this dimension has a very low importance rating, which suggests that subjects did not differentiate stories very highly on this dimension. The third dimension is virtually uninterpretable and its importance is very low, 0.09. Most of the error from the two dimensional solution may have been relegated to this third dimension, causing the reduction in stress found in moving from the 2 dimensional solution to the 3 dimensional solution.

The analysis of the accomplished subjects' data is slightly more difficult. It may be that the accomplished subjects had a more difficult task in sorting the stories into groups than either the expert or the naive subjects. Here the model explains 93.3 percent of the variance in the data—a remarkably good fit. Figure 2 displays the graph of dimension 1 versus 2. The first dimension has an importance of 0.50 and seems to differentiate between the regression and law of large number principles. However, this relationship is not as strong as in the expert data: one of the stories is mis-classified by this dimension. Figure 3, a graph of dimensions 1 versus 3, makes visible a rather unique interpretation. Namely, the four center points all display Fong, Krantz, and Nisbett's probabilistic domain, while the points outside that ellipse show the objective and subjective

domains. One entity, Story E, does not fit closely in the core ellipse, however, it is in the general area of the other four points. Also as Figure 4 shows, dimension 3 seems to separate the stories into 3 distinct groups, however, there seems to be no underlying similarities among those groups. In short, I can find no simple interpretation for dimensions 2 and 3, when each is examined alone. While dimension 1 seems to exhibit a polar symmetry, dimensions 2 and 3 do not clearly differentiate any aspects of the stories.

The data from the naive subjects shows a different pattern than either the expert or accomplished subjects. Even here, the model explains 92.9 percent of the variance in the data. The first dimension, with an importance of 0.45, distinguished between the academic and sports contexts rather well. Only 3 stories were mis-classified on the basis of this dimension and all of those stories had low stimulus weights. Figure 5, a graph of dimensions 1 and 2, clearly supports this interpretation. Also the second dimension, with an importance of 0.35, perfectly distinguishes between the principles of regression and the law of large numbers. The interpretation of the second dimension is clearly one of statistical principle. The third dimension shows no clear cut pattern among the stories and appears to be uninterpretable. Also its importance is very low, 0.13—indicating that subjects did not use this dimension.

Figure 1: Expert MDS Analysis
Dimension 1—Principle versus Dimension 2—Context
Importance of the dimension is indicated in parentheses beside the dimension.

Legend

<i>Context</i>	S	Sports			C	Class
<i>Principle</i>	R	Regression			L	Law of Large Numbers
<i>Domain</i>	P	Probabilistic	O	Objective	S	Subjective
<i>Stories</i>	A	Championship Selection Problem			I	Teaching Psychology
	B	Basketball Talent Scout			J	Stanbrook's Housing Office
	C	Timmy			K	David Chooses a College
	D	Harold—High School Football Coach			L	Howard—Junior High Teacher
	E	Baseball Statistics			M	Janet and the Gifted Students
	F	Competitive Ice Skating			N	Joe, Retiring Dean
	G	Rookie of the Year			O	The Class Selection Problem
	H	Slot Machines

Figure 2: Accomplished MDS Analysis
Dimension 1—Principle versus Dimension 2—Unknown
Importance of the dimension is indicated in parentheses beside the dimension.

Legend

<i>Context</i>	S	Sports	C	Class		
<i>Principle</i>	R	Regression	L	Law of Large Numbers		
<i>Domain</i>	P	Probabilistic	O	Objective	S	Subjective
<i>Stories</i>	A	Championship Selection Problem	I	Teaching Psychology		
	B	Basketball Talent Scout	J	Stanbrook's Housing Office		
	C	Timmy	K	David Chooses a College		
	D	Harold—High School Football Coach	L	Howard—Junior High Teacher		
	E	Baseball Statistics	M	Janet and the Gifted Students		
	F	Competitive Ice Skating	N	Joe, Retiring Dean		
	G	Rookie of the Year	O	The Class Selection Problem		
	H	Slot Machines

Figure 3: Accomplished MDS Analysis
Dimension 1—Principle versus Dimension 3—Unknown
Importance of the dimension is indicated in parentheses beside the dimension.

Legend

<i>Context</i>	S	Sports	C	Class
<i>Principle</i>	R	Regression	L	Law of Large Numbers
<i>Domain</i>	P	Probabilistic	O	Objective
			S	Subjective
<i>Stories</i>	A	Championship Selection Problem	I	Teaching Psychology
	B	Basketball Talent Scout	J	Stanbrook's Housing Office
	C	Timmy	K	David Chooses a College
	D	Harold—High School Football Coach	L	Howard—Junior High Teacher
	E	Baseball Statistics	M	Janet and the Gifted Students
	F	Competitive Ice Skating	N	Joe, Retiring Dean
	G	Rookie of the Year	O	The Class Selection Problem
	H	Slot Machines

Figure 4: Accomplished MDS Analysis
Dimension 2—Unknown versus Dimension 3—Unknown
Importance of the dimension is indicated in parentheses beside the dimension.

Legend

<i>Context</i>	S	Sports	C	Class
<i>Principle</i>	R	Regression	L	Law of Large Numbers
<i>Domain</i>	P	Probabilistic	O	Objective
			S	Subjective
<i>Stories</i>	A	Championship Selection Problem	I	Teaching Psychology
	B	Basketball Talent Scout	J	Stanbrook's Housing Office
	C	Timmy	K	David Chooses a College
	D	Harold—High School Football Coach	L	Howard—Junior High Teacher
	E	Baseball Statistics	M	Janet and the Gifted Students
	F	Competitive Ice Skating	N	Joe, Retiring Dean
	G	Rookie of the Year	O	The Class Selection Problem
	H	Slot Machines

Figure 5: Naive MDS Analysis

Dimension 1—Context versus Dimension 2—Principle

Importance of the dimension is indicated in parentheses beside the dimension.

Legend

<i>Context</i>	S	Sports	C	Class
<i>Principle</i>	R	Regression	L	Law of Large Numbers
<i>Domain</i>	P	Probabilistic	O	Objective
			S	Subjective
<i>Stories</i>	A	Championship Selection Problem	I	Teaching Psychology
	B	Basketball Talent Scout	J	Stanbrook's Housing Office
	C	Timmy	K	David Chooses a College
	D	Harold—High School Football Coach	L	Howard—Junior High Teacher
	E	Baseball Statistics	M	Janet and the Gifted Students
	F	Competitive Ice Skating	N	Joe, Retiring Dean
	G	Rookie of the Year	O	The Class Selection Problem
	H	Slot Machines		

Cluster analysis and network analysis were also conducted on the data and are reported in Appendix B. The results of these two procedures confirmed the basic results from the multidimensional analysis: namely, that statistical principle was the primary dimension the experts used in sorting the stories, while the naive subjects primarily sorted the data by context. In addition the experts' networks tended to be more highly interconnected and relations between stories were well defined, where the naive subjects' networks were rather loosely connected and did not show any clear pattern linking the stories.

Discussion

The results of the four analyses conducted on the data uncovered several interesting facets of the subjects' perceptions of the stories. First, the expert subjects tended to differentiate the stories based upon the statistical principle embodied in the story. Second, the naive subjects grouped the stories according to the context of the story. Both of these outcomes support the earlier hypothesis, that novices would sort problems according to surface structure and experts would sort according to deep structure. The interpretation of the accomplished subjects' data was more difficult because these subjects did not use any of the hypothesized variables consistently. Also, Fong et al.'s (1986) contention that the stories could be distinguished by the domain and structure variables was not supported.

All of the analyses conducted on the experts' data supported the same conclusion: the experts' primary means of distinguishing the stories was based on the statistical principle involved in the stories. The primary dimension uncovered by the multi-dimensional scaling was a dimension which differentiated the stories embodying the regression principle from stories encompassing the law of large numbers principle. Also, the correlational analysis revealed that there was a great deal of agreement among the experts. They all tended to use the same structure to describe the stories and there was high inter-rater reliability. These analyses strongly indicate that the experts differentiated the stories primarily based on the statistical principle involved.

The interpretation of the naive subjects data is slightly more complicated than that of the experts. The multi-dimensional scaling results indicated that these subjects sorted the stories based primarily on a context dimension—however they did not use that dimension consistently. It is interesting to note that the dimensions used by the naive subjects are the reverse of those used by the experts. The experts' primary dimension was statistical principle and their second dimension was context. The naive subjects primary dimension was context and their secondary dimension was statistical principle. (It may seem odd that naive subjects perceived the statistical principle dimension at all in the sortings. I believe this lends support to the notion that some statistical principles are used as heuristics. This may be one method of coping with the extraordinary amount of information to which people are exposed. Or it may be that naive subjects were not so naive.) Finally, the correlational analysis indicated very little reliability among the naive subjects themselves. Also the average correlation with the expert subjects was low—indicating little agreement between expert and naive subjects.

The interpretation of the accomplished subjects' data is at best confusing. These subjects display some of the attributes of the experts but also seem confused about how to apply the rules they have learned. The results of the multi-dimensional scaling analyses showed that these subjects used some of the same dimensions as the experts (their primary dimension was statistical principle). However, the second and third dimensions were uninterpretable. There was no consistent pattern in either of those dimensions. The subjects seem to be at an intermediate state in their knowledge of statistics, and while they seem to understand the general ideas they have a difficult time with unusual occurrences. However, this is reasonable given that these subjects are in the middle of learning about these concepts.

Most importantly, only one of the accomplished subjects' multi-dimensional scaling graphs showed any support for Fong et al.'s (1986) domain variable. All other analyses, especially for the expert and naive subjects, did not reveal any effect of either dimension (domain or structure) in the sorting of the problems. This indicates that subjects do not perceive stories in

different domains as being different. These findings do not support Fong, et al.'s. Subjects did perceive certain problem similarities, but not the ones defined by Fong, et al. (1986). Thus transfer across problems differing in domain type (probabilistic, objective, or subjective) may not indicate any generalization if context and principal are identical.

Experiment 2

Experiment 2 was similar to the second experiment conducted by Fong et al. (1986) with a more stringent control on training time, a different stimulus set, and a different testing procedure to reduce the potential for demand characteristics. The stimulus set used in this study was based on the results of Experiment 1 and Fong, et al's (1986) study. The stimulus set was representative of both the primary dimension along which naive subjects identified problems in statistical heuristics, context (sports or academic) and Fong, et al's (1986) domain (probabilistic, objective, and subjective) dimension. Designing the stimulus set around problems which naive subjects perceived differently provided a more thorough test of the transfer of statistical heuristics to new problems. An expanded testing procedure also provided more information on how and when subjects apply statistical heuristics by examining the problem of demand characteristics. Specifically, if subjects were using statistical heuristics when reasoning in the laboratory and not in everyday situations then the effects of the use of statistical heuristics should be examined more carefully.

Subjects were placed into one of four experimental groups—sports probabilistic, academic objective, combined training, and no training. Training time was equivalent for all groups. The first three groups were similar to Fong, Krantz, and Nisbett's full training condition, in that subjects received training on both rules and examples. However, for the first group—sports probabilistic—the examples came from problems which had a sports context and a probabilistic domain. Similarly, for the second group—academic objective—the examples came from problems which had an academic context and an objective domain. For the combined training group one example was from the academic objective set, while the other example was from the sports

probabilistic set. Finally, the no-training group completed a distractor task. Immediately after this training subjects were tested in accordance with the procedure used by Fong et al. (1986) in their first experiment.

Several problems from dimensions not included in training were included in the testing phase in order to assess the specificity of training and the transfer to different problem domains. Subjects were then told to return in two weeks for a post-test. No problems were used in both the training and the testing phase. Before the post-test, which was conducted with the same protocol as the first test, subjects were called and asked to answer questions pertaining to biographical information, statistical heuristics, and several distractors. These questions were disguised in the form of an attitude survey and all the problems on statistical heuristics had appropriate cover stories. This design allowed for a more conclusive test of the transfer of training to different problems and also served to control demand characteristics. If subjects were responding to the mere presentation of statistical heuristics then the attitude survey would reveal this problem.

This experiment shed light on the transfer of statistical training to novel statistical problems. In their study, Fong et al. (1986) defined probabilistic, objective, and subjective problems as different domains and they found, using these domains, that training generalized to different domains. However, if these problem domains are not perceived by subjects, then their interpretation of the results is suspect. This experiment should indicate whether or not training transfers to domains which subjects perceive as different.

Method

Subjects

The subjects were chosen from the the Rice University Psychology Subject Pool for the fall semester of 1989. A total of 48 subjects voluntarily participated in the experiment for partial fulfillment of course credit. The average age of the subjects was 19.5 years. The standard deviation was 3.5 years. Fifteen males (31% of the subjects) and thirty-three females (69% of the subjects) were tested. Approximately 50% of the sample had received no statistical training, while

47% had received 1 or 2 courses in statistics. One subject, 3% of the sample, had 3 courses in statistics. The subjects were unevenly distributed among the 5 basic academic majors granted by the university. Business, Natural Science, and Social Science major each contributed approximately 25% of the sample. The remaining 25% of the sample was evenly divided between Engineering and Humanities majors. In addition, 62% of the subjects were tested in 4 or 5 person groups, while the remaining 38% of the sample was tested in either 2 or 3 person groups. Complete demographic statistics for the sample are reported in Table F-1.

Design

The experimental design was a 4 (training condition) $\times$ 3 (test phase) $\times$ 4 (test problem) mixed design. Training condition was manipulated as a between subjects variable, while both test phase and test problem were manipulated as repeated measures variables. The four training conditions were Sports Probabilistic (SPT), Academic Objective (AOT), Combined Training (CT), and No Training (NT). The first three groups were similar to Fong et al's (1987) full training condition, in that subjects received training on both rules and examples. The three test phases were Test 1 (T1), Phone survey (P), and Test 2 (T2). The four test problems are described below. The order of presentation of the short stories was randomized and counter-balanced across subjects with the constraint that one of each type of test problem was to be included in Test 1 and Test 2. For the telephone survey the order of presentation of problems and distractor questions was randomized for each subject.

Stimuli

Four basic types of test problems and two types of training problems were included in each test phase described above. The problems were short stories which were based on the law of large numbers and written around the context (sports and academic) and domain (probabilistic, objective, and subjective) dimensions described earlier. The four test problem types were developed by crossing the levels of the context dimension with the first two levels of the domain dimension resulting in—sports probabilistic, sports objective, class probabilistic, and class

objective. Because Experiment 1 showed that subjects could not reliably differentiate the three levels of the domain variable, only the two levels which were the most different (probabilistic and objective) were used. This also reduced the complexity of the design. The two training problem types were sports probabilistic and academic objective. In addition two types of distractor problems were included in Test 1 and Test 2. One of the distractor problems was based on the regression principle—distractor regression—and the other—distractor other—had no unique dimension represented in its structure.

For each of the six kinds of problems—sports probabilistic (SP), sports objective (SO), academic probabilistic (AP), academic objective (AO), distractor regression (DR), and distractor other (DO)—2 different short stories were generated; one short story for each of the two test phases. (Four of these short stories were also used in Experiment 1.) Short questions for each of the four basic problem types—AO, AP, SO, SP—were generated for the telephone survey. (Short stories were considered unlikely material to be found in an attitude survey, therefore one sentence questions were used instead.) Statistical distractor questions were not used in the telephone survey because of time limitations, however other types distractor questions were used. Four additional short stories were written for the example training: two each of the sports probabilistic and academic objective problems types. Thus, a total of twenty problems (16 testing stories and 4 training stories) were written for this experiment. The stories used in training are included in Appendix C and the stories used in testing are included in Appendix D. A complete list of the dimensions corresponding to each problem is reported in Table D-1.

Reading grade level analyses (See Table 5) were conducted on the short stories to ensure that all stories had equivalent levels of cognitive complexity. This was intended to reduce the chance that subjects would be influenced by the overall difficulty of interpreting and analyzing the stories. Results indicate that the stories were similar in structure, vocabulary, and syntax. The average number of words per sentence ranged from 11.9 to 25.3, with a mean of 16.89 (SD = 4.50). The average number of one syllable words for the stories was 68.1%. The Dale-Chall,

Forecast, and Fry method were used to determine the reading grade levels of all stories. The average reading grade level for the Dale-Chall, Forecast, and Fry method across the 16 stories was 9-10, 9.21, and 8.47. The other four problems used in the phone survey were not written as short stories and thus were not included in this analysis.

Several analyses of variance were carried out on the 16 short stories used in Test 1 and Test 2. The Forecast reading grade level was used as the dependent measure in all ANOVA's, while the factors that the problems were written around and the different testing conditions were used as independent measures. The ANOVA's indicated that there were no significant differences in the reading grade levels between training problems and testing problems ($F(1,10) = 0.692$, $p = 0.4250$); between distractor problems, sports test problems, and academic test problems ($F(2,9) = 0.2220$, $p = 0.1645$); and on the context ($F(1,8) = 2.148$, $p = 0.1809$); and domain ($F(1,8) = 0.618$, $p = 0.4545$) factors. In addition, the reading grade levels of all the stories were within ± 2 standard deviations of the grand mean (using the Forecast method). The average reading grade level (Forecast method) of the stories used in the training phase of the study was 8.38, with a standard deviation of 0.95, while the average reading grade level of the test problems was 9.47 with a standard deviation of 0.73.

Procedure

Subjects were tested in small groups of two to five people. Each group was brought into the laboratory and randomly assigned to one of the four training conditions. At the beginning of the first testing session, subjects read and signed an informed consent form. The subjects were told,

This is an experiment on human reasoning. Today you will participate in a short training procedure and then you will read, analyze, and comment on the logic and reasoning involved in 6 short stories. In two weeks you will return to this laboratory and you read, analyze, and comment on the logic and reasoning involved in 6 more short stories.

Subjects in the SP, AO, and Combined Training conditions spent the next 15 minutes reading a paper on the law of large numbers and the basic ideas of populations, samples, and distributions. Then these subjects read the first example training problem and were instructed to think about how the concepts in the paper applied to the example. The subjects were given a few minutes to digest the story, then the experimenter discussed the application of the law of large numbers to the population and sample distributions of the example training problem. Subjects read the second example problem and then discussed it with the experimenter following the same protocol as for the first example. The discussion of both example problems lasted approximately fifteen minutes. Appendix C contains both the rules training paper and the examples used in discussing the law of large numbers with subjects.

The two example problems were varied for each of the three training conditions—SP, AO, Combined Training. The subjects given the SP training reviewed and discussed example training problems written around the probabilistic domain and involving a sports context. The AO training groups read and discussed training problems written around the objective domain and involving an academic context. The groups which received Combined Training read and discussed one example from the sports probabilistic set of stories and one from the academic objective set of stories. The No Training or control condition was given a distractor task involving the identification of warnings. The training for all four conditions lasted approximately 30 minutes.

After the training all subjects were told,

You will now read, analyze, and comment on the logic and reasoning involved in 6 short stories. After you have completed this task you are free to go. Please remember that you must return in two weeks to complete this experiment in order to receive credit.

Then subjects were given the first set of six problems. They were told to analyze and write comments about each problem. This took approximately 45 minutes. Subjects were then told to return in two weeks for the second test. This concluded the first session.

Table 5

Reading Level Analyses of Stimulus Materials for the Second Experiment

Story Identifier	Mean Words per Sentences	Percentage of One Syllable Words	Reading Grade Levels		
			Dale-Chall	Forecast	Fry
Training Problems					
AO1	18.4	69.0	9-10	9.0	10.0
AO5	17.5	66.3	9-10	9.5	9.0
SP4	15.0	65.5	9-10	9.6	8.0
SP5	16.6	78.9	7-8	7.4	6.0
Mean	16.9	69.9	9-10 [*]	8.9	8.3
Standard Deviation	15	62	7-8 to 9-10 ^b	09	1.7
Testing Problems					
AO2	25.3	56.9	College	11.0	College
AO4	17.6	65.9	9-10	9.6	9.0
AP1	15.1	69.8	7-8	8.9	7.0
AP2	14.4	60.7	9-10	10.4	8.0
SO1	18.0	67.6	9-10	9.3	9.0
SO2	22.6	68.1	11-12	9.2	10.0
SP1	19.4	67.6	11-12	9.3	9.0
SP2	17.1	78.1	9-10	7.5	6.0
DO1	16.8	69.0	9-10	9.0	9.0
DO2	12.5	64.0	7-8	9.9	8.0
DR2	15.4	64.3	7-8	9.8	9.0
DR3	17.0	67.6	9-10	9.3	9.0
Mean	17.6	66.6	9-10 ^a	9.4	8.8
Standard Deviation	35	52	7-8 to College ^b	09	1.7
Grand Mean	17.4	67.5	9-10 [*]	9.3	8.7
Grand Standard Deviation	3.1	5.4	7-7 to College ^b	0.9	1.7

^a Mode^b Range

One week after the first session subjects were called and asked to participate in an attitude survey about Rice University. Twenty-six subjects (54% of the sample) were contacted and agreed to participate in the study. The complete protocol and the questions used are found in Appendix E. Subjects did not know that this telephone survey was part of the experimental design. If subjects chose to participate in the telephone survey they were asked a total of 14 questions. Ten of the questions were distractors which dealt with university administration, the library, the food plan, and other facets of undergraduate life. Four of the questions were written along SP, AO, AP, and SO dimensions. These 14 questions were presented in a random order. Subjects' answers to the questions were taped and later coded. If subjects chose not to participate they were thanked for their time and the telephone survey was concluded. At least four attempts were made to contact each subject.

Subjects returned to the laboratory two weeks after the initial session, approximately one week after the telephone survey, and completed the experiment. The subjects were seated in the laboratory and told, "You will now read, analyze, and comment on the logic and reasoning involved in these 6 stories." Then the subjects read, analyzed and made comments about the remaining six short stories. This test was completed in 45 minutes. After the testing, subjects were asked to fill out biographical information blanks and to state the amount of statistical experience they had received and their familiarity with sports. At the end of the second session subjects were debriefed.

Analysis

The comments generated by each subject for every story were rated by three independent raters. The raters were told to rate each story on a scale of 1 to 3. A '1' indicated that the subjects response was completely deterministic. A '2' indicated that there was some evidence of statistical thinking, but the subject's answer was incomplete or confused. A '3' indicated a good statistical answer, that is the subject indicated that sample size was a concern and that as the number of samples increased so did the accuracy of the estimate. All three raters

agreed on 344 of the 576 (59.7%) “answers”. Two of the three raters agreed on 563 of the 576 (97.7%) answers. The average inter-rater correlation was 0.773. These results indicated that the three raters coded the stories on similar attributes and that the agreement between raters was acceptable. In order to analyze the data the ratings from each rater were reduced. The final rating was based on the rating given by at least 2 of the 3 raters. If all three raters disagreed then the answer was rated as a ‘2’. The final rating was used as the dependent measure in all subsequent analyses.

Results

Two basic analyses were conducted: item analysis and Analysis of Variance. The item analysis was conducted to determine whether problem difficulty would unduly influence the later ANOVA's. Then the data were analyzed with the Analysis of Variance (ANOVA) procedure as implemented in SAS version 5. Two ANOVA's were conducted. The first was a three factor mixed design, 4 training conditions x 2 test phases X 6 test problems. This analysis included data from all subjects who participated in the experiment and was designed to examine the effect of transfer of training on the different test problems. Because the response rate for the phone survey was quite low (54%), another analysis was conducted using only those 26 subjects who participated in the phone survey. This analysis was designed to examine the demand characteristics which might be present in the laboratory testing procedure. For this second analysis the phone survey was added as a third test phase and the two distractor problems were dropped from test problem factor, since the distractors were not included in the phone survey. Thus, the design for the second analysis was 4 training conditions x 3 test phases x 4 test problems.

Item Analysis

Since story difficulty could unduly influence the results of later ANOVA's, item analyses were conducted on the 16 stories used in Experiment 2. The item statistics are reported in Table 6. In general the stories showed acceptable levels of difficulty. The distractor questions—DO and DR—were given statistical answers much less frequently than the other questions—AO, AP, SO,

and SP. This was expected since the distractors do not have any statistical content. Also, this was used as a manipulation check. If subjects were not analyzing the problems as they were told they would respond to every problem, including the distractors, with a statistical answer. Since the mean ratings of the distractor questions were lower than the mean ratings of the other questions, it appears that subjects did complete the task as instructed.

The item analysis revealed rather large differences among the items. The range of means went from 1.0, for the SOP question, to 2.6, for the SO1 question. Both academic probabilistic questions, AP1 and AP2, had very high mean ratings. This indicates that there is quite a lot of variance among the items and, further, that the variance in the items is related to the effect of individual test questions. When further analyses of the effects of test problems are conducted, these differences will have to be taken into account.

Table 6.

Item Analyses for the 16 Test Questions Used as Stimuli in Experiment 2

Item Name	N	Mean	Standard Deviation	Min	Max	Skewness	Kurtosis	Percentage of Answers Rated as		
								1	2	3
AO2	48	2.00	0.82	1	3	0	-1.53	33.3	33.3	33.3
AO4	48	1.62	0.82	1	3	0.81	-1.00	58.3	20.8	20.8
AOP	26	1.04	0.20	1	2	5.09	26.00	96.2	3.8	0.0
AP1	48	2.46	0.77	1	3	-1.02	-0.51	16.7	20.8	62.5
AP2	48	2.35	0.73	1	3	-0.67	-0.81	14.6	35.4	50.0
APP	26	1.12	0.33	1	2	2.56	4.91	88.5	11.5	0.0
DO1	48	1.04	0.29	1	3	6.93	48.00	97.9	0.0	2.1
DO2	48	1.42	0.71	1	3	1.43	0.58	70.8	16.7	12.5
DR1	48	1.21	0.58	1	3	2.66	5.67	87.5	4.2	8.3
DR3	48	1.27	0.49	1	3	1.60	1.76	75.0	22.9	2.1
SO1	48	2.65	0.61	1	3	-1.52	1.33	6.3	22.9	70.8
SO2	48	2.08	0.85	1	3	-0.16	-1.59	31.3	29.2	39.6
SOP	26	1.00	0.00	1	1	n/a	n/a	100.0	0.0	0.0
SP1	48	2.17	0.78	1	3	-0.31	-1.28	22.9	37.5	39.6
SP2	48	1.77	0.92	1	3	0.48	-1.70	56.3	10.4	33.3
SPP	26	1.04	0.20	1	2	5.10	26.00	96.2	3.8	0

Legend*Context (Column 1)*

S Sports

A Academic

D Distractor

Domain (Column 2)

P Probabilistic

O Objective

R Regression

Test Phase (Column 3)

1 Test 1

2 Test 2

P Phone

Test Problem Analysis

The first analysis was conducted to examine the relationship between transfer of training and problem domain. This analysis did not include the phone survey. The main effect of training condition (see Figure 6) was significant ($F(3,44) = 3.75, p = 0.0175$) and a subsequent Newman-Keuls range test indicated that the Sports Probabilistic (SP) group performed significantly better than the Control (None) group ($p < 0.01$). This suggests that the sports probabilistic training was more beneficial to subjects than either the academic objective or combined training which were not different from the control condition. It seems that the academic objective (AOT) and combined training (CT) groups did not learn the concepts any better than the no training (NT) group did. Also the main effect of test phase (see Figure 7) was significant ($F(1,44) = 6.01, p = 0.0183$). Examination of the means of Test 1 (Mean = 1.92) and Test 2 (Mean = 1.76) indicated that subjects' performance became less statistical over time. This is not surprising since subjects' performance typically worsens over time.

Figure 6: Main Effect of Training Condition for the Test Problem Analysis.

Figure 7: Main Effect of Test Phase for the Test Problem Analysis.

Finally, the main effect of test problem (see Figure 8) was significant ($E(5,220) = 82.85$, $p < 0.0001$). A Subsequent Newman-Keuls range test revealed that performance on the distractor problems—DO and DR—was significantly worse ($p < 0.01$) than performance on the content problems—AO, AP, SO, and SP. This indicates that the subjects were not responding to every problem with a statistical answer and that they were attempting to analyze the stories and apply the statistical training appropriately. Also performance on the academic probabilistic (AP) and sports objective (SO) problems was significantly ($p < 0.01$) better than the academic objective (AO) and sports probabilistic (SP) problems.

The reasons for this difference are not immediately clear. One might contend that the AP and SO problems were simply easier for subjects. The item analysis supports this idea. However, the reading grade level analyses were performed to demonstrate that the stories were similar on most other extraneous dimensions. One explanation consistent with these facts is that the reading grade level analyses were not sophisticated enough to detect other factors which could affect statistical reasoning.

Figure 8: Main Effect of Test Problem for the Test Problem Analysis.

The only significant interaction (see Figure 9) was training condition by test problem ($F(15,220) = 1.79$, $p = 0.0366$). The 4 training groups (AO, SP, Combined, None) performed differently on the 6 test problems (AO, AP, SO, SP, DO, and DR). At this point a transfer of training analysis was conducted. It may seem strange to conduct an analysis of transfer of training for training groups whose learning did not differ from the no-training group. However, there might have been differences in the test problems which averaged out over the entire analysis. Thus an examination of the transfer of training was conducted for all training groups.

In order to test the a priori hypothesis that training would transfer across the domain dimension and not the context dimension, a careful examination of the interaction of training condition and test problem was carried out. In order to eliminate the effect of test problems from the analysis of transfer of training, difference scores were calculated by subtracting the mean of the no-training group for each problem from each subject's score in the three training conditions. The averaged ratings of the difference scores are referred to as mean differences in all further analyses.

Figure 9: Training Condition by Test Problem Interaction for the Test Problem Analysis.

This point is clarified in Table 7. The first column of the table refers to three of the four training conditions. The second column includes the test problems which were based on different dimensions from the training examples. For example, in the first line of the table (the academic objective training group) the mean difference for the sports probabilistic (SP) question is shown in the second column. This means that the SP problem differed on two dimensions (in this case context and domain) from the training example. The third column lists the test problems which differed from the training problems by only one dimension. The fourth column includes the test problems which were based on the same dimensions as the training examples. From this analysis subjects tended to do better on test problems which were based on the same dimensions rather than test problems which differed from training problems by one dimension and test problems which had no common dimensions. However, there were no statistically significant differences among the mean differences scores of the problems for subjects in the academic objective (AOT) or combined training (CT) groups. There was however a difference between the

sports probabilistic (SP) and sports objective (SO) problem for the Sports probabilistic (SPT) training group. Training transferred less to the sports objective problem than the sports probabilistic problem (Newmann-Keuls, $p < 0.01$). This difference is probably due to the fact that the sports probabilistic training transferred so well to sports probabilistic training and not because of any deficiency in the sports objective problem. In fact, training in other groups transferred to the sports objective group at approximately the same level as the SPT group.

Table 7

Training Conditions and Mean Differences for Test Problems Which Were the Same as the Training Examples and Those Which Differed from the Training Examples by One and Two Dimensions.

Training Condition	Different on Both Dimensions		Different on One Dimension		Same on Both Dimensions	
	AO	SP	SO	AP	AO	SP
Academic Objective (AOT)	–	0.32	0.41	0.34	0.64	–
Sports Probabilistic (SPT)	0.75	–	0.41	0.65	–	0.95
Combined (CT)			0.36	0.36	0.75	0.54

An approach to assessing the a priori hypothesis that training would transfer across the domain dimension and not the context dimension, is to examine the mean differences of the test problems (see Table 8). The first column of the table lists 3 of the training conditions. The second column refers to the mean differences of the academic and sports test problems when collapsing across the domain dimension, whereas the third column refers to the mean differences of the objective and probabilistic problems when collapsing across the context dimension. The t-tests below each pair of mean differences indicate whether they are significantly different. If the difference is significant then training did not transfer across that dimension for that training group. This analysis indicates training transferred across both dimensions for all training groups.

Table 8

Mean Differences in Statistical Rating Averaged Across Same versus Different Test Problem Dimensions for Academic Objective (AOT), Sports Probabilistic (SPT), and Combined (CT) Training.

Training Condition	Transfer Across Context Dimension		Transfer Across Domain Dimension	
	Academic	Sports	Objective	Probabilistic
Academic Objective (AOT)	0.49 $t(12) = 1.27, n.s.$	0.37	0.53 $t(12) = 1.80, n.s.$	0.33
Sports Probabilistic (SPT)	0.70 $t(10) = 1.04, n.s.$	0.68	0.58 $t(10) = -1.57, n.s.$	0.80
Combined (CT)	0.56 $t(10) = 1.00, n.s.$	0.45	0.56 $t(10) = 1.45, n.s.$	0.45

Phone Survey Analysis

The Phone survey was conducted to determine whether demand characteristics associated with post tests 1 and 2—both of which were conducted in the laboratory—could account for transfer effects. If subjects produced fewer statistical answers for problems on the telephone survey than they did for the problems on Test 1 or Test 2 then the demand characteristics of the laboratory setting maybe have contributed to the transfer effects. This hypothesis appeared to be supported. Subjects produced fewer statistical answers to the problems on the telephone survey (see Figure 10) than to problems on either of the post tests ($E(2,44) = 71.38, p = 0.0001$). A follow-up Newman-Keuls range test indicated that answers on the telephone survey were significantly less statistical than either post test ($p < 0.01$) and that answers from Test 1 and Test 2 were not different ($p > 0.05$).

Figure 10: Main Effect of Test Phase for the Phone Survey Analysis.

There was a significant ($F(6,44) = 2.41$, $p = 0.0422$) test phase by training condition interaction (see Figure 11). This interaction was primarily due to the fact that the effects of training seen in the earlier analysis were greatly reduced in the phone survey. It is interesting to note that analysis of the simple effects indicates that the training conditions responded with statistical answers which were reliably different on the first test ($p < 0.01$), but not on the second test ($p > 0.05$). However, when data from all subjects are analyzed without phone survey results there is no training by test phase interaction. This discrepancy is probably due to sampling differences.

Figure 11: Test Phase by Training Condition Interaction for the Phone Survey Analysis.

The test phase by test problem interaction (see Figure 12) was also significant ($E(6,132) = 2.4$, $p = 0.0438$). The interaction indicated that the effect of problem was reduced in the phone survey. The test problems were different at Test 1 ($E(3,66) = 3.924$, $p = 0.012$) and Test 2 ($E(3,66) = 7.915$, $p = 0.001$). Test problems were not different on the phone survey ($E(3,66) = 31.197$, $p = 0.318$).

Figure 12: Test Phase by Test Problem Interaction for the Phone Survey

Analysis.

As in the previous analysis the main effect of training condition (see Figure 13) was marginally significant ($F(3,22) = 2.84$, $p = 0.0615$). Again the subjects in the SP group responded with more statistical answers to the questions than did subjects in the control, AOT, or CT groups. Finally, the main effect of test problem (see Figure 14) was significant ($F(3,66) = 15.41$, $p = 0.0001$). Again, as in the first analysis, subjects responded to the SO and AP problems with significantly more statistical answers than to the the AO and SP problems (Newman-Keuls, $p < 0.01$).

Figure 13: Main Effect of Training Condition for the Phone Survey Analysis.

Figure 14: Main Effect of Test Problem for the Phone Survey Analysis.

Discussion

The first analysis was conducted to examine whether formal statistical training transfers to different problems when the problem domains are subjectively defined. The answer appears to be 'yes', with the exception of the SP training transferring so well to the SP problem. Neither the academic objective nor the sports probabilistic training groups generated different levels of statistical answers for the 4 content questions (AO, AP, SO, and SP). Thus, it appears that training does transfer to different domains when the domains are subjectively defined. There was success in replicating Fong et al.'s (1986) claim that training does transfer between the objective and probabilistic problems. Training transferred across both dimensions, context and domain, for all training groups.

One provocative issue discovered in the analyses was that Sports Probabilistic training yielded more statistical answers than Academic Objective or Combined training. There are two possible causes for this finding. First, although randomly assigned, the subjects in the Academic Objective training group had less statistical experience than subjects in the Sports Probabilistic and Combined training groups. Of the subjects in the AOT group 78.6% had no statistical experience, whereas 25% of the subjects in the CT and SPT had no statistical training. It could be that the SPT and CT subjects' previous statistical knowledge allowed them to grasp the concepts addressed in the training more easily than could the AOT subjects.

Second, sports problems are generally more concrete and may lend themselves to statistical analysis more easily than do the academic problems. In fact, Nisbett claims that sports "culture" encourages statistical reasoning. For instance, it is much easier to think of the Rookie of the Year problem in terms of statistical percentages than it is the professor who is trying to decide which textbook to use. Also, during discussion of the problems in the training sessions, students appeared to understand the sports questions more easily than the academic questions. This may also explain why the AO training group did not learn the statistical content of the training session very well.

In addition, the academic problems may be more difficult than the sports problems because academic problems do not lend themselves to statistical analysis as easily as sports problems do. And the probabilistic problems may be easier than the objective problems because they explicitly cue the subjects to the fact there is an underlying statistical explanation. Thus you would expect subjects to perform less well on problems which contained two conceptually difficult levels of the dimensions (i.e., academic and objective) than all other problems which contained the easier levels of the dimensions (either probabilistic or sports). Consequently, the poor performance of the AO training group might be due to a general difficulty with academic objective problems. This also helps to explain the poor performance of the CT group. Since the CT group received both a sports probabilistic (SP) and an academic objective (AO) problem, it may be that the AO problem confused the subjects.

Another curiosity that this experiment revealed is the fact that SO and AP problems were better in two test phases (Test 1 and Test 2). The item analysis indicates that the SO and AP problems, averaged across all tests, were simply easier than the other questions. Again, this is in line with the idea that sports and probabilistic dimensions are easier for subjects to grasp.

Another problem with this experiment concerns the low average statistical score. (The mean across all questions was 1.64.). Although the distribution of scores was acceptable (58.7% of the questions were given a rating of '1', 14.7% were given a rating of '2', and 26.7% were given a rating of '3') it seems that subjects either understood the statistical nature of the questions or they mostly did not. Since the questions may not have been well understood by some of the subjects the effects of transfer of training may have been moderated. It is probable that the amount of training and its content were not sufficient for subjects to gain a full understanding of the principles of the law of large numbers and sample size and therefore they may not have been able to adequately apply them. In fact, this seems quite reasonable. The primary implication of this finding is the restriction of range in the criterion variable. This indicates that there is reduced power and thus the chances of finding a significant interaction are reduced.

The effect of demand characteristics on the two post tests was demonstrated with the telephone survey. The subjects responded with fewer statistical answers to the telephone survey than to the two other post tests, indicating that there was a significant laboratory bias. This seems to indicate that there was an effect on statistical reasoning simply by bringing the subjects in to a laboratory and asking them to reason. It also casts doubt on the degree to which the results of previous laboratory studies generalize to behavior outside the laboratory.

Poor performance on the phone survey can be explained in terms of demand characteristics, but alternately this may be due to the fact that 1) the test questions on the phone survey were strange, or 2) the coding of subjects' answers was difficult. Specifically, the questions which subjects were asked had to be very short, as with the other questions in a phone survey, yet still had to reflect the basic nature of probabilistic or objective questions. It proved to be very difficult to write questions which were both sensible and accurate. For instance, the SP question read, "How do you feel the performance of the Rice University football team will compare with past years, even though the order of the teams they play against is essentially random?" The last phrase of the statement was included to reflect the randomizing device which must be in a probabilistic problem. However, it does not make a great deal of sense when the question is considered. Of course, it could be argued that this made the randomizing condition even more salient and thus subjects should have grasped the statistical content of the questions more easily.

The second issue involves the coding of the answers from the phone survey. Essentially any answer which reflected an appreciation of the law of large numbers would have been scored as statistical. For instance, if a subject had said, "I don't know, but with more games we can make a more accurate prediction," in response to the SP problem it would have earned a very high statistical rating. However, if a subject had said, "They are a really bad team," in response to the question it would have been scored as deterministic. Other answers which were coded as deterministic responses were, "They lose every year," and "The coach has not had time to develop a good team." These last two problems may reflect an implicit understanding of the law of

large numbers, since the Rice football team does not very often win games, however there was not enough information given to adequately evaluate the subjects' knowledge. Although both of these problems reduce the generalizability of the results, they do not preclude the fact that transfer to the telephone problems was nonexistent. Thus, demand characteristics of the laboratory were very much in effect and should be taken into account in future research.

General Discussion

The first experiment was conducted to determine which dimensions are salient to subjects in defining problem domains. Three separate dimensions—context, principle, and domain—were combined to form 15 short stories. Subjects then completed a sorting task on the stories. The results indicated that naive subjects sorted the stories based on the context dimension, while experts sorted the stories based on the content dimension. Neither group used the domain dimension as specified by Fong et al. (1986). They found that training subjects on certain problem domains resulted in transfer of training to all domains.

The second study attempted to expand upon Fong et al's (1986) original study by pairing Fong et al's (1986) domain dimension with a stimulus dimension which was subjectively defined and by examining the demand characteristics of the laboratory. The results indicated that training transferred along both the domain dimension and the context dimension. This result does not clearly support either the rule or example learning hypotheses.

The transfer found in the second experiment was affected by demands of the laboratory testing procedure. Specifically, subjects responded to questions in the phone survey with answers which were significantly less statistical than answers in the two test sessions. There was a significant effect of statistical reasoning induced simply by laboratory testing. This has implications which are quite strong. Primarily, since most previous research on reasoning has been conducted in the laboratory, the effect of training may have been overstated. This is of particular concern with Fong et al's (1986) study.

There were limitations in these experiments. One problem with this research is that it was conducted primarily on college students. Although this poses no theoretical restraints on the generalizability of the study, the results should be "handled with care." In the first study the sample size of the experts was small ($n=6$). With a greater sample size the results would be substantially more reliable. For the second study, a more thoroughly developed set of stimulus materials would have helped. A thorough item analysis conducted before the tests and phone survey would have eliminated some of the problems associated with relative item difficulties found in this study and would have enhanced the validity of the study by ensuring adequate reliabilities for each of the test phases.

So, the primary question remains. What are the cognitive structures which accompany learning when training transfers to different domains? The two basic philosophies which have evolved posit two opposite views: Namely, 1) that abstract rules are learned, or 2) that concrete problem instances are learned. If abstract rules are learned then training should transfer to all novel problems. If concrete examples are learned then transfer of training should occur to the extent that they share features which map onto learned examples. Because training did transfer to all novel problems, it is possible that an abstract rule which encompassed the entire set of problems was learned. However, the training tended to transfer best when the test problems and the training problems were based on the same dimensions, and less well when training problems differed from testing problems by one dimension. For instance, for the sports probabilistic training group, training appears to transfer to the SP problem best and slightly less well to the AP and SO problem. To the extent that the problems shared common features there was greater transfer of training. This pattern of results is characteristic of learning examples.

One possibility is that both rule and example learning occur simultaneously, and that in trying to create a situation where either rules or examples solely enhance training, an artificial dichotomy has been created. In fact, the most fruitful lines of research may emerge from considering the aspects of the training which enhance learning. Consider the work of Nitsch

(1977), who found that training on definitions and examples produced greater transfer than training on definitions alone. Specifically, the definitions training group studied the definition of a word created explicitly for this study, while the definitions and examples group studied the definitions of the word plus examples of the word's usage in sentences. The definitions alone condition bears striking resemblance to rules training. Since training time was controlled, this seems to indicate that the combination of definitions and examples, or possibly rules and examples, provides better transfer than either technique separately.

Further research in the area could concentrate on the effect of the environment on reasoning. Since there is a tremendous difference between reasoning in the laboratory and in a more natural environment, the attitude survey might be exploited to ascertain the magnitude of these differences and the patterns of subjects' use of statistical heuristics. Another line of research in this area might concentrate on the possible cognitive operations which underlie reasoning, such as analogical transfer. Also, the basic parameters which determine the aspects of training which do transfer could be investigated. It also remains to examine what aspects of the training process encourage transfer to different domains and what aspects discourage transfer. This type of research would have tremendous impact upon both training and teaching programs.

In summary, the first hypothesis, concerning transfer of training to different problems when the differences were subjectively defined, was supported. In the laboratory studies, training transferred across all subjectively defined boundaries. This result does not clearly support either the view that rules or examples are learned. The second hypothesis dealt with the transfer of training to different problems (or even the same type of problem) when the potential for demand characteristics was reduced. Specifically, training does not transfer well at all (even to the same type of problem) when the contextual effects of the questions and the environment are removed.

References

- Adelson, B. (1981). Problem solving and the development of abstract categories in programming languages. Memory & Cognition, 9, 422-433.
- Agnoli, F., & Krantz, D. H. (1989). Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy. Cognitive Psychology, 21, 515-550.
- Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition. Psychometrika, 35, 238-319.
- Caylor, J. S., Sticht, T. G., Fox, L. C., & Ford, J. P. (1973). Methodologies for Determining Reading Requirements of Military Occupational Specialties. Technical Report No. 73-5, HUMRO Western Division, Presidio of Monterey, CA: Human Resources Research Organization.
- Chase, W.G., & Simon, H.A. (1973). Perception in chess. Cognitive Psychology, 5, 55-81.
- Chi, T.H.C., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. Cognitive Science, 2, 121-132.
- Cooke, N. J., & Schvaneveldt, R. W. (1988). Effects of computer programming experience on network representations of abstract programming concepts. International Journal of Man-Machine Studies, 29, 407-427.
- Dale, E., & Chall, J. S. (1984). A formula for predicting readability. Educational Research Bulletin, 27, 11-20.
- de Groot, A. D. (1965). Thought and Choice in Chess. The Hague: Mouton.
- Eddy, D.M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press.
- Egan, D. E., & Schwartz, B. J. (1979). Chunking in the recall of symbolic drawings. Memory & Cognition, 7, 149-158.
- Fisk, A. D., & Gallini, J. K. (1989). Training consistent components of tasks: Developing an instructional system based on automatic/controlled processing principles. Human Factors, 31, 453-463.
- Fong, D.H., Krantz, D.H., Nisbett, R.E. (1986). The effects of statistical training on thinking about everyday problems. Cognitive Psychology, 18, 253-292.
- Fry, E. (1968). A readability formula that saves time. Journal of Reading, 11, 513-516, 575-578.
- Hayes, J. R., & Simon, H. A. (1977). Psychological differences among problem isomorphs. In N. J. Castellan, D. B. Pisoni, and G. R. Potts (Eds.), Cognitive Theory, Volume 2. Hillsdale, NJ: Erlbaum.

- Howell, C.H., & Cooke, N.J. (1989). Training the human information processor: A review of cognitive models. In I. L. Goldstein and Associates (Eds.), Training and Development in Organizations (pp. 121-182). San Francisco: Jossey Bass, Inc.
- Isaac, P. D., & Poor, P. D. S. (1974). On the determination of appropriate dimensionality in data with error. Psychometrika, 39, 91-109.
- Kahneman, D. & Tversky, A. (1972). Subjective probability: A judgment of representativeness. Cognitive Psychology, 3, 430-454.
- Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80, 237-251.
- Kaiser, M.K., Jonides, J. , & Alexander, J. (1986). Intuitive reasoning about abstract and familiar physics problems. Memory & Cognition, 14, 308-312.
- Kruskal, J. B. & Carroll, J. D. (1969). Geometric models and badness-of-fit functions, pp. 689-670 in P. R. Krishnaiah (Ed.) Multivariate Analysis, Vol. 2. New York: Academic Press.
- Kruskal, J. B. & Wish, M. (1978). Multidimensional Scaling. Beverly Hills, CA: Sage Publications.
- Lehman, D.R., Lempert, R.O., & Nisbett, R.E. (1988). The effects of graduate training on reasoning: Formal discipline and thinking about everyday-life events. American Psychologist, 43, 431-442.
- McDonald, J. E. & Schvaneveldt, R. W. (1987). The Application of User Interface Knowledge to Interface Design. In Raymonde Guindon (Ed.), Cognitive Science and its Applications for Human Computer Interaction (pp. 289-338). Hillsdale, NJ: Lawrence Erlbaum Associates.
- McDonald, J. E., Dayton, T. & McDonald, D. R. (1986). Adapting Menu Layout to Tasks. Memorandum in Computer and Cognitive Science (MCCS-86-78). Las Cruces, NM: New Mexico State University Computing Research Laboratory.
- Newell, A., & Simon, H. A. (1972). Human Problem Solving. Englewood Cliffs, NJ: Prentice-Hall.
- Nisbett, R.E. & Ross, L. (1980). Human Inference: Strategies and Shortcomings of social judgement. Englewood Cliffs, NJ.:Prentice-Hall.
- Nisbett, R.E., Fong, G.T., Lehman, D. R., Cheng, P. W. (1987). Teaching Reasoning. Science, 238, 625-631.
- Nisbett, R.E., Krantz, D.H., Jepson, K., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. Psychological Review, 90, 339-363.
- Nitsch, K. E. (1977). Structuring decontextualized forms of knowledge. Ph.D. Dissertation, Vanderbilt University.
- Piaget, J., & Inhelder, B. (1975). The origin of the idea of chance in children. New York: Norton. (Originally published in 1951.)
- Reitman, J. S. (1976). Skilled perception in go: Deducing memory structures from inter-response times. Cognitive Psychology, 8, 336-356.

- Rosenberg, S., & Kim, P. (1975). The method of sorting as a data gathering procedure in multivariate research. Multivariate Behavioral Research, 10, 489-502.
- Rosenberg, S., Nelson, C., & Vivikananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. Journal of Personality and Social Psychology, 9, 283-294.
- SAS Institute Inc. (1985). SAS Users Guide: Statistics, Version 5 Edition. Cary, NC: SAS Institute.
- SAS Institute Inc. (1989). SUGI Supplemental Library Users Guide, Version 5 Edition. Cary, NC: SAS Institute.
- Schoenfeld, A.H., & Herrmann, D.J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 484-494.
- Schvaneveldt, R.W., Durso, F.T., & Dearholt, D.W. (1985). Pathfinder: Scaling with network structures. Memorandum in Computer and Cognitive Science (MCCS-85-9). Las Cruces, NM: New Mexico State University Computing Research Laboratory.
- Simon, H. A., & Chase, J. R. (1973). Skill In chess. American Scientist, 61, 394-403.
- Simon, H. A., & Hayes, J. R. (1976). The understanding process: Problem isomorphs. Cognitive Psychology, 8, 165-190.
- Sokal, R. R. & Michener, C. D. (1958). A statistical model for evaluating systematic relationships. University of Kansas Science Bulletin, 38, 1409-1438.
- Thorndike, E.L. (1906). Principles of Teaching. New York: A.G. Seiler.
- Tversky, A. & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5, 207-232.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristic and biases. Science, 185, 1124-1131.

APPENDIX A

Stimuli and Reading Grade Level Analyses for the Stories Used in Experiment 1.

Context: Sports

Principle: Law of Large Numbers

Domain: Probabilistic

Structure: 1

Source: Fong 7

Identification: B

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. He flips a coin to decide which game he should see first. The player he observes is generally excellent. He repeatedly makes plays worthy of the best professional players. However, in one of the games, with his team behind by 2 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 to 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 has (b) "a tendency to misplay under extreme pressure." Does the thinking behind either conclusion have any weaknesses?

DESCRIPTIVES:	GRADE LEVELS:
Number of sentences.....11	Coleman-Liau 10.2
Number of words.....205	Dale-Chall 9-10 (44 words not in list)
Number of syllables.....304	Flesch..... 9.8 (REI is 61.1)
Number of letters961	Flesch-Kincaid..... 9.2
Number of words of 6 or more letters 73	Forecast..... 9.3
Number of words of 7 or more letters 45	Fry 9.0
Number of 1 syllable words138	Gunning-Fog.....12.1
Number of words of 3 or more syllables..... 24	Holmquist 9.4
Percentage of 1 syllable words.....67.3	Lix..... 9.0 (Lix index is 40)
Percentage of words of 3 or more syllables11.7	Powers..... 6.0
Average number of words per sentence18.6	Raygor12.0
Average number of letters per word..... 4.7	Rix 9.0 (Rix index is 4.09)
Average number of syllables per word 1.5	McLaughlin-Smog.....11.6

Context: Class

Principle: Law of Large Numbers

Domain: Probabilistic

Structure 1

Source: Fong 1

Identification: J

At Stanbrook University, the Housing Office determines which of the 10,000 students enrolled will be allowed to live on campus the following year. At Stanbrook, the dormitory facilities are excellent, so there is always great demand for on-campus housing. Unfortunately, there are only enough on-campus spaces for 5000 students. The Housing Office determines who will get to live on campus by having a Housing Draw every year: every student picks a number out of a box over a 3-day period. These numbers range from 1 to 10,000. If the number is 5000 or under, the student gets to live on campus. If the number is over 5000, the student will not be able to live on campus.

On the first day of the draw, Joe talks to five people who have picked a number. Of these, four people got low numbers. Because of this, Joe suspects that the numbers in the box were not properly mixed, and that the early numbers are more favorable. He rushes over to the Housing Draw and picks a number. He gets a low number. He later talks to four people who drew their numbers on the second or third day of the draw. Three got high numbers. Joe says to himself, "I'm glad that I picked when I did, because it looks like I was right that the numbers were not properly mixed."

What do you think of Joe's reasoning? Explain.

DESCRIPTIVES:	GRADE LEVELS:
Number of sentences..... 16	Coleman-Liau 6.8
Number of words..... 242	Dale-Chall 7-8 (32 words not in list)
Number of syllables..... 338	Flesch..... 7.7 (REI is 73.1)
Number of letters1011	Flesch-Kincaid..... 6.8
Number of words of 6 or more letters 65	Forecast..... 8.9
Number of words of 7 or more letters39	Fry 7.0
Number of 1 syllable words169	Gunning-Fog..... 8.7
Number of words of 3 or more syllables16	Holmquist..... 7.9
Percentage of 1 syllable words..... 69.8	Lix..... 6.0 (Lix index is 31)
Percentage of words of 3 or more syllables 6.6	Powers..... 5.3
Average number of words per sentence..... 15.1	Raygor 7.0
Average number of letters per word..... 4.2	Rix 7.0 (Rix index is 2.44)
Average number of syllables per word 1.4	McLaughlin-Smog..... 8.8

Context: Sports

Principle: Law of Large Numbers

Domain: Subjective

Structure 1

Source: Fong 13

Identification: C

Gerald M. had a three year old son, Timmy. He told a friend: "You know, I've never been much for sports, and I think Timmy will turn out the same. A couple of weeks ago, an older neighbor boy was tossing a ball to him, and he could catch it and throw it all right, but he just didn't seem interested in it. Then the other day, some kids his age were kicking a little soccer ball around. Timmy could do it as well as the others, but he lost interest very quickly and started playing with some toy cars while the other kids went on kicking the soccer ball around for another 20 or 30 minutes."

Do you agree with Gerald's reasoning that Timmy is likely not to care much for sports?

Why or why not?

DESCRIPTIVES:

Number of sentences..... 9
 Number of words.....137
 Number of syllables..... 176
 Number of letters 537
 Number of words of 6 or more letters 26
 Number of words of 7 or more letters 13
 Number of 1 syllable words 104
 Number of words of 3 or more syllables 5
 Percentage of 1 syllable words..... 75.9
 Percentage of words of 3 or more syllables 3.6
 Average number of words per sentence 15.2
 Average number of letters per word..... 3.9
 Average number of syllables per word 1.3

GRADE LEVELS:

Coleman-Liau 5.3
 Dale-Chall 5-6 (10 words not in list)
 Flesch..... 6.9 (REI is 81.4)
 Flesch-Kincaid..... 5.5
 Forecast 7.9
 Fry 6.0
 Gunning-Fog..... 7.5
 Holmquist..... 5.4
 Lix..... 5.0 (Lix index is 24)
 Powers..... 4.8
 Raygor 6.0
 Rix 5.0 (Rix index is 1.44)
 McLaughlin-Smog..... 7.4

Context: ? Principle: Law of Large Numbers

Domain: Probabilistic Structure 2

Source: Fong 2 Identification: H

After one of their football games, Keith and several of his friends decided to go out for the night. They spent much of the evening talking about the game and trying decide where they made their mistakes and what they could do prevent them from happening again. On the way home Keith stopped at a gas station. There, in one corner of the gas station, were two slot machines. Keith had heard about slot machines, but had never actually seen one. He went over to the slot machines and looked at them, trying to figure out how they worked. An old man who was sitting close to the machines spoke to Keith. "There ain't no winning system for slot machines. It's all luck. You just put in a coin, pull the lever and hope that you'll win. But let me tell you this: some machines are easier to lose on than others. That's because the owners can change the mechanism of the slots so that some of them will be more likely to make you lose. See those two slot machines there? The one on the left gives you about an even chance of winning, but the one on the right is fixed so that you'll lose much more often than you'll win. Take it from me--I've played them for years." The old man then got up and walked out of the gas station.

Keith was by now very intrigued by the two slot machines, so he played the machine on the left for a couple of minutes. He lost almost twice as often as he won. "Humph," Keith said to himself. "The man said that there was an even chance of winning at that machine on the left. He's obviously wrong." Keith then tried the machine on the right for a couple of minutes and ended up winning more often than he lost. Keith concluded that the man was wrong about the chances of winning on the two slot machines. He concluded that the opposite was true--that the slot machine on the right was more favorable to the player than the machine on the left.

Comment on Keith's conclusion and his reasoning. Do you agree? Explain your answer.

DESCRIPTIVES:

Number of sentences.....	27	GRADE LEVELS:	
Number of words.....	372	Coleman-Liau	6.3
Number of syllables.....	473	Dale-Chall.....	5-6 (36 words not in list)
Number of letters.....	1535	Flesch.....	6.7 (REI is 82.8)
Number of words of 6 or more letters	78	Flesch-Kincaid.....	4.8
Number of words of 7 or more letters	51	Forecast.....	7.6
Number of 1 syllable words	290	Fry	5.0
Number of words of 3 or more syllables	16	Gunning-Fog.....	7.2
Percentage of 1 syllable words.....	78.0	Holmquist.....	8.3
Percentage of words of 3 or more syllables	4.3	Lix.....	5.0 (Lix index is 27)
Average number of words per sentence.....	13.8	Powers.....	4.7
Average number of letters per word.....	4.1	Raygor	6.0
Average number of syllables per word	1.3	Rix	6.0 (Rix index is 1.89)
		McLaughlin-Smog.....	7.5

Context: Sports

Principle: Law of Large Numbers

Domain: Objective

Structure: ?

Source: Nisbett & Krantz p.356

Identification: A

Two sports fans are arguing over which sport--baseball or football-- has the best (most accurate) play-off system. Charlie says that the Super Bowl is the best way of determining the world champion because, according to him, "the seven games of the world series are all played in the home cities of the two teams, whereas the Super Bowl is usually played in a neutral city. Since you want all factors not related to the game to be equal for a championship, then the Super Bowl is the better way to determine the world championship." Which procedure do you think is a better way to determine the world champion--World Series or Super Bowl? Why?

DESCRIPTIVES:

Number of sentences..... 5
 Number of words..... 115
 Number of syllables..... 168
 Number of letters 512
 Number of words of 6 or more letters 30
 Number of words of 7 or more letters 21
 Number of 1 syllable words 79
 Number of words of 3 or more syllables..... 13
 Percentage of 1 syllable words..... 68.7
 Percentage of words of 3 or more syllables .. 11.3
 Average number of words per sentence..... 23.0
 Average number of letters per word..... 4.5
 Average number of syllables per word 1.5

GRADE LEVELS:

Coleman-Liau 9.1
 Dale-Chall.....11-12 (25 words not in list)
 Flesch.....11.0 (REI is 56.6)
 Flesch-Kincaid.....10.6
 Forecast..... 9.1
 Fry 9.0
 Gunning-Fog.....13.7
 Holmquist..... 7.5
 Lix..... 9.0 (Lix index is 41)
 Powers..... 6.2
 Raygor 9.0
 Rix 9.0 (Rix index is 4.2)
 McLaughlin-Smog....12.3

Context: Class	Principle: Law of Large Numbers
Domain: Subjective	Structure: ?
Source: Nisbett & Krantz p.356	Identification: O

It is the first week of the winter term. Henry has signed up for five classes, but plans to take only four. Three of these classes he knows he wants, so he must decide between the other two. Both are on subjects interesting to him. The student course evaluations indicate that Course A is better taught. However, he attended the first meeting of both classes this week, and found Course B's session more enjoyable. Which class should he take and why?

DESCRIPTIVES:	GRADE LEVELS:
Number of sentences..... 7	Coleman-Liau 7.7
Number of words..... 83	Dale-Chall 5-6 (6 words not in list)
Number of syllables..... 113	Flesch..... 7.4 (REI is 76.3)
Number of letters 367	Flesch-Kincaid..... 5.1
Number of words of 6 or more letters 23	Forecast..... 7.9
Number of words of 7 or more letters 14	Fry 6.0
Number of 1 syllable words 63	Gunning-Fog..... 7.7
Number of words of 3 or more syllables..... 6	Holmquist..... 4.7
Percentage of 1 syllable words..... 75.9	Lix..... 6.0 (Lix index is 28)
Percentage of words of 3 or more syllables 7.2	Powers..... 4.9
Average number of words per sentence..... 11.9	Raygor 7.0
Average number of letters per word..... 4.4	Rix 6.0 (Rix index is 2)
Average number of syllables per word 1.4	McLaughlin-Smog..... 8.4

Context: Sports

Principle: Regression

Domain: Objective

Structure: ?

Source: Nisbett & Krantz p.359

Identification: G

In general the major league baseball player who wins Rookie of the Year does not perform as well in his second year. This is clear in major league baseball in the past ten years. In the American League, 8 rookies of the year have done worse in their second year; only 2 have done better. In the National League, the rookie of the year has done worse the second year 9 times out of 10.

Why do you suppose the Rookie of the Year tends not to do as well his second year?

DESCRIPTIVES:

Number of sentences.....	5	GRADE LEVELS:	
Number of words.....	93	Coleman-Liau	5.9
Number of syllables.....	124	Dale-Chall	7-8 (12 words not in list)
Number of letters	369	Flesch	7.2 (REI is 78)
Number of words of 6 or more letters	22	Flesch-Kincaid.....	7.4
Number of words of 7 or more letters	9	Forecast.....	8.5
Number of 1 syllable words	67	Fry	7.0
Number of words of 3 or more syllables	4	Gunning-Fog.....	9.2
Percentage of 1 syllable words.....	72.0	Holmquist.....	5.8
Percentage of words of 3 or more syllables	4.3	Lix.....	6.0 (Lix index is 28)
Average number of words per sentence.....	18.6	Powers.....	5.3
Average number of letters per word.....	4.0	Raygor	7.0
Average number of syllables per word	1.3	Rix	6.0 (Rix index is 1.8)
		McLaughlin-Smog.....	8.2

Context: Sports

Principle: Regression

Domain: Probabilistic

Structure: 3

Source: Fong 3

Identification: E

Bert H. has a job tabulating statistics for a baseball team. Bert has notice that one batter hits a home-run 80% of the time. Of 100 at-bats each day, usually about 70 to 85 are home-runs. Bert has noticed that on some days, all of the first 10 hits are home-runs. However, Bert has also noticed that on such days, the overall number of home-runs is usually not much better for the day as a whole than on days when the first 10 hits are not home-runs.

Why do you suppose the number of home-runs is usually not much better on days where the first groups of hits yield home-runs than on other days?

DESCRIPTIVES:

Number of sentences.....	7	GRADE LEVELS:	
Number of words.....	114	Coleman-Liau	6.4
Number of syllables.....	160	Dale-Chall	9-10 (19 words not in list)
Number of letters	465	Flesch.....	7.8 (REI is 71.9)
Number of words of 6 or more letters	26	Flesch-Kincaid.....	7.3
Number of words of 7 or more letters	20	Forecast	8.0
Number of 1 syllable words	86	Fry	7.0
Number of words of 3 or more syllables.....	14	Gunning-Fog.....	11.4
Percentage of 1 syllable words.....	75.4	Holmquist	6.4
Percentage of words of 3 or more syllables ...	12.3	Lix.....	7.0 (Lix index is 33)
Average number of words per sentence.....	16.3	Powers.....	5.5
Average number of letters per word.....	4.1	Raygor	7.0
Average number of syllables per word	1.4	Rix	7.0 (Rix index is 2.86)
		McLaughlin-Smog....	11.2

Context: Sports

Principle: Regression

Domain: Subjective

Structure: 3

Source: Fong 15

Identification: F

Janice is the recruiting director for a competitive ice skating team. She says the following of her experiences: "There is big turnover of the athletes here, and each year we hire 15-20 new skaters. Some of these people show themselves to be unusually warm and friendly in the first few days. One might set up a party for the whole group. Another might help a fellow skater learn a new routine. I find though that, over the long haul, these skaters turn out to be not much more concerned and caring than the others. What happens, I think, is that they can't remain open and vulnerable without paying a heavy emotional price. They usually continue to be considerate but they build up a shell."

Comment on Janice's reasoning. Do you think it is likely that she correctly identifies the skaters who are unusually warm and friendly? Do you agree it is likely that most of the ones who are unusually warm at first later build up a shell to protect themselves emotionally?

DESCRIPTIVES:

Number of sentences.....	12	Coleman-Liau	8.5
Number of words.....	173	Dale-Chall	7-8 (25 words not in list)
Number of syllables	259	Flesch	8.9 (REI is 65.3)
Number of letters	774	Flesch-Kincaid	7.7
Number of words of 6 or more letters	46	Forecast	8.8
Number of words of 7 or more letters	35	Fry	8.0
Number of 1 syllable words	122	Gunning-Fog	10.4
Number of words of 3 or more syllables	20	Holmquist	7.0
Percentage of 1 syllable words.....	70.5	Lix	7.0 (Lix index is 34)
Percentage of words of 3 or more syllables ...	11.6	Powers	5.7
Average number of words per sentence	14.4	Raygor	7.0
Average number of letters per word.....	4.5	Rix	7.0 (Rix index is 2.92)
Average number of syllables per word	1.5	McLaughlin-Smog.....	10.5

Context: Class

Principle: Law of Large Numbers

Domain: Subjective

Structure: 2

Source: Fong 14

Identification: K

David L. was a senior in high school on the east coast who was planning to go to college. he had compiled an excellent record in high school and has been admitted to his two top choices: a small liberal arts college and an Ivy League university. The two schools were about equal in prestige and were equally costly. Both were located in attractive east cost cities, about equally distant from his home town. David had several older friends who were attending the liberal arts college and several who were attending the Ivy League university. They were all excellent students like himself and had interests that were similar to his. His friends at the liberal arts college ail reported that they liked the place very much and that they found it very stimulating. The friends at the ivy League university reported that they had many complaints on both personal and social grounds and on educational grounds. David initially he thought he would go to the liberal arts college. However, he decided to visit both schools himself for the day. He did not like what he saw at the private liberal arts college: several people whom he met seemed cold and impersonal; a professor he met with briefly seemed abrupt and uninterested in him; and he did not like the "feel" of the campus. He did like what he saw at the Ivy League university: several of the people he met seemed like vital, enthusiastic, pleasant people; he met with two different professors who took a personal interest in him; and he came away with a very pleasant feeling about the campus.

What school should David L. choose, and why? Try to analyze the arguments on both sides, and explain which side is stronger.

DESCRIPTIVES:

Number of sentences..... 14
 Number of words..... 290
 Number of syllables 444
 Number of letters1340
 Number of words of 6 or more letters 95
 Number of words of 7 or more letters 72
 Number of 1 syllable words 195
 Number of words of 3 or more syllables 43
 Percentage of 1 syllable words..... 67.2
 Percentage of words of 3 or more syllables ... 14.8
 Average number of words per sentence 20.7
 Average number of letters per word..... 4.6
 Average number of syllables per word 1.5

GRADE LEVELS:

Coleman-Liau 9.9
 Dale-Chall 9-10 (52 words not in list)
 Flesch.....10.3 (REI is 58.9)
 Flesch-Kincaid.....10.6
 Forecast..... 9.3
 Fry10.0
 Gunning-Fog.....14.2
 Holmquist.....10.4
 Lix.....10.0 (Lix index is 45)
 Powers..... 6.4
 Raygor11.0
 Rix10.0 (Rix index is 5.14)
 McLaughlin-Smog....13.1

Context: Class

Principle: Regression

Domain: Objective

Structure: 3

Source: Fong 9

Identification: L

Howard was a teacher in junior high school in a community known for truancy and delinquency problems among its youth. Howard says of his experiences: "Usually, in a class of 35 or so kids, 2 or 3 will pull some pretty bad stunts in the first week—they'll skip a day of class, get into a scuffle with another kid, or some such thing. When that kind of thing happens, I play it down and try to avoid calling the class' attention to it. Usually, these kids turn out to be no worse than the others. By the end of the term you'll find they haven't pulled any more stunts than the others have." Howard reasons as follows: "Some of these kids are headed toward a delinquent pattern of behavior. When they find out nobody is very impressed, they tend to settle down."

Comment on Howard's reasoning:

a) Do you agree that it is likely that the students who pull a "pretty bad stunt in the first week" are "headed toward a delinquent pattern of behavior?"

b) Do you agree that it is likely that the students who initially pull a "pretty bad stunt" turn out to be no worse than the others because they find no one is impressed with their behavior?

DESCRIPTIVES:

Number of sentences..... 12
 Number of words..... 213
 Number of syllables..... 296
 Number of letters 891
 Number of words of 6 or more letters 57
 Number of words of 7 or more letters 33
 Number of 1 syllable words 157
 Number of words of 3 or more syllables..... 18
 Percentage of 1 syllable words..... 73.7
 Percentage of words of 3 or more syllables 8.5
 Average number of words per sentence..... 17.8
 Average number of letters per word..... 4.2
 Average number of syllables per word 1.4

GRADE LEVELS:

Coleman-Liau 7.1
 Dale-Chall 7-8 (33 words not in list)
 Flesch..... 8.0 (REI is 70.3)
 Flesch-Kincaid..... 7.7
 Forecast..... 8.3
 Fry 7.0
 Gunning-Fog.....10.5
 Holmquist..... 8.1
 Lix..... 7.0 (Lix index is 33)
 Powers..... 5.5
 Raygor 8.0
 Rix 7.0 (Rix index is 2.75)
 McLaughlin-Smog.....10.1

Context: Sports

Principle: Regression

Domain: Objective

Structure: 3

Source: Nisbett p. 354

Identification: D

Harold is the coach for a high school football team. One of his jobs is selecting new members of the varsity team. He says the following of his experience: "Every year we add 10-20 new boys to the team on the basis of their performance at the try-out practice. Usually the staff and I are extremely excited about the potential of two or three of these kids—one who throws several brilliant passes or another who kicks several field goals from a remarkable distance. Unfortunately, most of these kids turn out to be only somewhat better than the rest."

Why do you suppose that the coach usually has to revise downward his opinion of players that he originally thought were brilliant?

DESCRIPTIVES:

Number of sentences.....	7	Coleman-Liau	9.0
Number of words.....	122	Dale-Chall	9-10 (22 words not in list)
Number of syllables.....	187	Flesch.....	9.5 (REI is 62.3)
Number of letters	550	Flesch-Kincaid.....	9.3
Number of words of 6 or more letters	35	Forecast	8.8
Number of words of 7 or more letters	29	Fry	9.0
Number of 1 syllable words	86	Gunning-Fog.....	13.5
Number of words of 3 or more syllables.....	20	Holmquist	6.8
Percentage of 1 syllable words.....	70.5	Lix.....	9.0 (Lix index is 41)
Percentage of words of 3 or more syllables ...	16.4	Powers.....	6.1
Average number of words per sentence.....	17.4	Raygor	9.0
Average number of letters per word.....	4.5	Rix	9.0 (Rix index is 4.14)
Average number of syllables per word	1.5	McLaughlin-Smog....	12.8

Context: Class	Principle: Law of Large Numbers
Domain: Objective	Structure: ?
Source: Experimenter	Identification: I

Mark is a teacher at a small mid-western college. During his first year of teaching, he selected a new textbook for his introductory class, An Introduction to Psychology. He likes this new book very much, however, some of his colleagues at a much larger university have been using both his book and a different text, Psychology for the Ages, in their introductory psychology classes. His colleagues report that, Psychology for the Ages, is much easier for the students to understand and that the students seem to like the book better.

Should Mark switch to using the new book, Psychology for the Ages, or should he continue to use the one he likes best?

DESCRIPTIVES:

Number of sentences.....	6
Number of words.....	113
Number of syllables.....	175
Number of letters	526
Number of words of 6 or more letters	32
Number of words of 7 or more letters	24
Number of 1 syllable words	81
Number of words of 3 or more syllables	18
Percentage of 1 syllable words.....	71.7
Percentage of words of 3 or more syllables ...	15.9
Average number of words per sentence	18.8
Average number of letters per word.....	4.7
Average number of syllables per word	1.5

GRADE LEVELS:

Coleman-Liau	10.0
Dale-Chall	9-10 (19 words not in list)
Flesch.....	9.8 (REI is 60.9)
Flesch-Kincaid.....	10.0
Forecast	8.6
Fry	10.0
Gunning-Fog.....	13.9
Holmquist.....	6.6
Lix.....	9.0 (Lix index is 40)
Powers.....	6.3
Raygor	8.0
Rix	9.0 (Rix index is 4)
McLaughlin-Smog....	13.0

Context: Class	Principle: Regression
Domain: Probabilistic	Structure: ?
Source: Experimenter	Identification: M

Janet gave a test to one of her elementary school classes, in order to select some members of the class for admission to a gifted students program. Four children from her class were selected to participate in this program. These four children were then randomly assigned to different groups. Each group was given specialized training for one hour each day after school. Several weeks later all the children were re-tested on a similar test. Janet was surprised to discover that three of the four students who were selected to participate in the gifted program did worse on the second test than on the first test. Janet reasoned that these two children were not paying attention to the test since they had already been selected for the gifted program and therefore their scores were lower.

Comment on Janet's reasoning. Do you think that her explanation is valid? Are there any other explanations which are plausible?

DESCRIPTIVES:		GRADE LEVELS:	
Number of sentences.....	10	Coleman-Liau	11.0
Number of words.....	154	Dale-Chall	9-10 (26 words not in list)
Number of syllables.....	236	Flesch.....	9.1 (REI is 64.3)
Number of letters	752	Flesch-Kincaid.....	8.5
Number of words of 6 or more letters	49	Forecast	9.8
Number of words of 7 or more letters	39	Fry	9.0
Number of 1 syllable words	99	Gunning-Fog.....	11.6
Number of words of 3 or more syllables.....	21	Holmquist.....	7.2
Percentage of 1 syllable words.....	64.3	Lix.....	9.0 (Lix index is 40)
Percentage of words of 3 or more syllables ...	13.6	Powers.....	6.0
Average number of words per sentence.....	15.4	Raygor	9.0
Average number of letters per word.....	4.9	Rix	9.0 (Rix index is 3.9)
Average number of syllables per word	1.5	McLaughlin-Smog....	11.4

Context: Class	Principle: Regression
Domain: Subjective	Structure: ?
Source: Fong 15	Identification: N

Joe is the retiring Dean of a large social sciences school. In his exit interview he reports the following of his experiences at the university: "One of the school's biggest problems is the turnover rate among new professors. Each year we hire between 5 and 10 new educators. A few of these men and women are very ambitious—they seem eager and enthusiastic about their new positions. As an example, one of them wrote a new textbook during her first year. While another played a major part in overhauling the graduate and undergraduate curricula. But, I usually find that after a several years, these professors are not much more energetic and eager than the others. I think that they put so much effort in at the beginning of their careers that they get burned out very quickly. Usually they continue to do research and contribute to the program at a normal rate, but nothing like what their first year pace would have suggested."

Comment on Joe's reasoning. Do you think it is likely that he correctly identifies the professors who are eager and ambitious? Do you agree it is likely that most of the ones who are very eager at first later become burned out?

DESCRIPTIVES:		GRADE LEVELS:	
Number of sentences.....	12	Coleman-Liau	9.4
Number of words.....	204	Dale-Chall	9-10 (40 words not in list)
Number of syllables.....	313	Flesch.....	9.5 (REI is 62.7)
Number of letters	933	Flesch-Kincaid.....	9.1
Number of words of 6 or more letters	56	Forecast.....	9.3
Number of words of 7 or more letters	44	Fry	9.0
Number of 1 syllable words	138	Gunning-Fog.....	12.9
Number of words of 3 or more syllables.....	31	Holmquist	8.9
Percentage of 1 syllable words.....	67.6	Lix.....	8.0 (Lix index is 38)
Percentage of words of 3 or more syllables ...	15.2	Powers.....	6.1
Average number of words per sentence.....	17.0	Raygor	8.0
Average number of letters per word.....	4.6	Rix	8.0 (Rix index is 3.67)
Average number of syllables per word	1.5	McLaughlin-Smog....	12.3

APPENDIX B

Cluster and Network Analysis Results for the Experiment 1.

Cluster Analysis

Cluster analyses were performed for each of the three expertise sub-groups. The analyses were performed on the sum of all co-occurrence matrices for subjects in a given expertise sub-group. Specifically, the cluster analysis method chosen was the average cluster linkage method developed by Sokal and Michener (1958). According to SAS (1985), the method computes the distances based on the average distance between pairs of observations and tends to combine clusters which have low variances. The program also computes a squared multiple correlation, R^2 , which is defined as $1 - (P_G/T)$, where P_G is similar to the sums of squares explained for all clusters and T is the total sums of squares.

The interpretation of the cluster analysis of the expert data is fairly simple (see Figure B-1). Two main groups were found: one representing the statistical principle of the law of large numbers and the other signifying the regression principle. These two clusters explain 80% of the variance in the expert data. The regression stories seemed to be more tightly connected, by virtue of their lower cluster distances, than the stories concerning law of large numbers.

The cluster analysis of the accomplished sub-groups data is shown in Figure B-2. The cluster diagram describes four basic groups, and explains 81% of the variance in the data. The first cluster, reading from the bottom up, is composed of stories which espoused the law of large numbers. The second cluster primarily contains stories with an academic context—one sports context story is included in this group. The third cluster is comprised entirely of sports regression stories. The fourth cluster encompasses stories with an academic context and the law of large numbers principle.

The data for the naive sub-group is enigmatic. There seem to be five basic clusters (see Figure B-3), which explain 81% of the variance in the data. The first cluster is composed of a single entity—story A. The second cluster is composed of stories which incorporated both the academic context and the statistical principle the law of large numbers. The third cluster is a mix of

both contexts—academic and sports, and both statistical principles—law of large numbers and regression. However, the majority of the stories in the cluster incorporated an academic context or the regression principle. The fourth cluster includes three stories which all concern the sports context and the principle of regression. The fifth cluster contains two law of large numbers stories. These five clusters seem to characterize the data quite well, but their interpretation in terms of context, domain, principle, or structure is puzzling.

Network Analysis

A network analysis of the subjects data was also conducted. Again, separate analyses were done for each expertise sub-group—expert, accomplished, and naive. The analyses were completed using the Pathfinder algorithm developed by Schvaneveldt, Durso, & Dearholt (1985). The general goal of the Pathfinder algorithm as explained by its authors is, “given the MAX network (which is defined by the data), a link is included in the Pathfinder link-weighted network if and only if the link is a minimum-length path between the pairs of nodes connected by the link. The length of the path is computed as a function of the weights associated with the links in the path, and a path may consist of any number links.” As McDonald & Schvaneveldt (1987) later explain, “Pathfinder. . . is based on an idea that a link is present in the output network if and only if that link is a minimum weight path between the nodes connected by the link in the (complete) network corresponding to input distances. Links are given weights corresponding to the pair-wise distances given as input, and the weight of a path is a function of the weights on the links in the path.” In general, the algorithm requires two parameters r and q , where r is the power in the Minkowski r -metric and q is the maximum number of links in paths examined in building the network. For all analyses reported below r equal infinity and q equal $n-1$. These settings allow for unique network structures with ordinal level data.

The pathfinder analysis of the expert subjects data revealed a very clear pattern which can be seen in Figure B-4. All of the regression principle stories were connected to all others, with the exception of Story E. This dense, star-like pattern indicates that these stories were highly

related. In contrast, the stories which embodied the principle of the law of large numbers were connected to each other linearly—by single lines. The expert subjects obviously saw these stories as being related. However, the relations between the stories were not as dense or interconnected as were the regression stories.

The data from the accomplished subjects are difficult to interpret (see Figure B-5). There is one cluster of law of large number stories. The backbone of the diagram is composed primarily of sports regression problems. However, two law of large number stories are connected at the end of the chain. At the bottom of the figure is a fairly dense cluster of mainly regression problems. Some of these stories are connected to several other stories in the group indicating a richer cognitive structure. However, there are some apparent inconsistencies. Two stories in that cluster concerned the law of large numbers and all six stories are evenly divided between sports and academic contexts.

The naive pathfinder network is shown in Figure B-6. The backbone of the structure encompasses stories with a sports context. At the last core node two groups of academic regression stories branch off and one these branches is connected to sports law of large numbers problems—obviously there was some confusion among these naive subjects as to how the stories were related. The top group of stories basically embodied the law of large numbers principle. These six stories are rather highly interconnected and the majority of them include academic contexts. This may indicate that subjects have some basic understanding of the law of large numbers.

Figure B-1: Expert Cluster Analysis

Legend*Context* S Sports

C Class

Principle R Regression

L Law of Large Numbers

Domain P Probabilistic O Objective

S Subjective

Stories A Championship Selection Problem
 B Basketball Talent Scout
 C Timmy
 D Harold—High School Football Coach
 E Baseball Statistics
 F Competitive Ice Skating
 G Rookie of the Year
 H Slot Machines

I Teaching Psychology
 J Stanbrook's Housing Office
 K David Chooses a College
 L Howard—Junior High Teacher
 M Janet and the Gifted Students
 N Joe, Retiring Dean
 O The Class Selection Problem

Legend

Context S Sports

C Class

Principle R Regression

L Law of Large Numbers

Domain P Probabilistic O Objective

S Subjective

Stories A Championship Selection Problem
 B Basketball Talent Scout
 C Timmy
 D Harold—High School Football Coach
 E Baseball Statistics
 F Competitive Ice Skating
 G Rookie of the Year
 H Slot Machines

I Teaching Psychology
 J Stanbrook's Housing Office
 K David Chooses a College
 L Howard—Junior High Teacher
 M Janet and the Gifted Students
 N Joe, Retiring Dean
 O The Class Selection Problem

Legend

C Class

L Law of Large Numbers

S Subjective

- I Teaching Psychology
- J Stanbrook's Housing Office
- K David Chooses a College
- L Howard—Junior High Teacher
- M Janet and the Gifted Students
- N Joe, Retiring Dean
- O The Class Selection Problem

Figure B-4: Expert Network Analysis (Pathfinder Algorithm)

Link weights are indicated by small numbers immediately beside the link. All unlabeled links are weighted zero.

Legend

<i>Context</i>	S	Sports	C	Class
<i>Principle</i>	R	Regression	L	Law of Large Numbers
<i>Domain</i>	P	Probabilistic	O	Objective
			S	Subjective
<i>Stories</i>	A	Championship Selection Problem	I	Teaching Psychology
	B	Basketball Talent Scout	J	Stanbrook's Housing Office
	C	Timmy	K	David Chooses a College
	D	Harold—High School Football Coach	L	Howard—Junior High Teacher
	E	Baseball Statistics	M	Janet and the Gifted Students
	F	Competitive Ice Skating	N	Joe, Retiring Dean
	G	Rookie of the Year	O	The Class Selection Problem
	H	Slot Machines

Figure B-5: Accomplished Network Analysis (Pathfinder Algorithm)
Link weights are indicated by small numbers immediately beside the link.

Legend

<i>Context</i>	S	Sports	C	Class
<i>Principle</i>	R	Regression	L	Law of Large Numbers
<i>Domain</i>	P	Probabilistic	O	Objective
<i>Stories</i>	A	Championship Selection Problem	I	Teaching Psychology
	B	Basketball Talent Scout	J	Stanbrook's Housing Office
	C	Timmy	K	David Chooses a College
	D	Harold—High School Football Coach	L	Howard—Junior High Teacher
	E	Baseball Statistics	M	Janet and the Gifted Students
	F	Competitive Ice Skating	N	Joe, Retiring Dean
	G	Rookie of the Year	O	The Class Selection Problem
	H	Slot Machines

Figure B-6: Naive Network Analysis (Pathfinder Algorithm)

Link weights are indicated by small numbers immediately beside the link.

Legend

<i>Context</i>	S	Sports	C	Class
<i>Principle</i>	R	Regression	L	Law of Large Numbers
<i>Domain</i>	P	Probabilistic	O	Objective
			S	Subjective
<i>Stories</i>	A	Championship Selection Problem	I	Teaching Psychology
	B	Basketball Talent Scout	J	Stanbrook's Housing Office
	C	Timmy	K	David Chooses a College
	D	Harold—High School Football Coach	L	Howard—Junior High Teacher
	E	Baseball Statistics	M	Janet and the Gifted Students
	F	Competitive Ice Skating	N	Joe, Retiring Dean
	G	Rookie of the Year	O	The Class Selection Problem
	H	Slot Machines		

APPENDIX C

Training Manual and Examples Used in Experiment 2.

INTRODUCTION

We are interested in studying how people go about explaining and predicting events under conditions of very limited information about the events. It is important to study how people explain and predict under these conditions because they frequently occur in the real world. Indeed, we often have to make important decisions based on such explanations and predictions, 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 is particularly helpful is called the Law of Large Numbers. Other related concepts which are helpful in understanding and explaining events are a population, a population distribution, a sample, and a sample distribution.

For the next few minutes we would like you to read the following training session and study the examples provided very carefully.

Populations and the concepts of sampling

Populations

In psychology the term population is used to mean the entire set of possible units for observation. These potential units can be real or hypothetical sets of people, plants, or animals. Populations can be very small or very large. For instance, the population of humans would consist of all humans now living on the earth. On the other hand, the population of professors at Rice University who have taught introductory psychology classes would include Bill Howell, Henry Roediger, John Brelsford, Sandra Carpenter, Sarah Burnett, Lila Laux, and Ken Laughery.

Population distributions

A population distribution describes both the range of values and the number of units at each value which would result if every member of the population were sampled. For example, our

population might be the gumballs in a gumball machine and we might be interested in the distribution of colored gumballs in the machine. So, let's assume that there are 10 gumballs in the machine and their colors are as follows: 1 is yellow, 2 are red, 3 are blue, 3 are orange, and 1 is white. Then the possible colors would be yellow, red, blue, orange, and white. And the number of gumballs with a certain color is 1 yellow, 2 red, 3 blue, 3 orange, and 1 white. Another way to represent the population distribution is by a graph. This is presented below.

The concept of a population distribution is used to describe the entire pool of units which are of interest in a particular study. In some sense, the population distribution tells the whole story about whatever you are studying. Consider the following: you are studying the effect of a multiple choice test format on introductory psychology students in the United States. If you tested all the students in introductory psychology courses on a multiple choice format test and then scored all the tests, that data would represent a population distribution. It would tell you the range of test scores and the number of individuals who achieved a certain score. This would represent the entire "story" for that group of students.

However, this would be very costly to administer and score and would probably take more time than you have to complete the study. In order to remedy this problem, we introduce the notion of a sample.

Samples

A sample is basically any subset of a population. It can contain as many or as few elements as you like with two preconditions: it must have at least one element from the population and can not contain all the elements in the population. The idea of a random sample is a very important idea for most scientists. A random sample simply means that the elements selected from the population were picked with no particular order in mind. They were selected at random from the population. This has a very important implication for the Law of Large Numbers which we will discuss in a minute. There are many different ways to generate a random sample. These include, mixing all the units up, perhaps in a jar, and drawing a sample from the middle of the jar, using a random number table, or using a random number generator function on a computer.

These random samples allow scientists to study a small subset of a population and then make inferences about the population based on the results of tests conducted on the sample. This is a very powerful tool in both the social and natural sciences.

Sample distributions

A sample distribution is very much like a population distribution. It describes both the range of values and the number of units at each value which would result if every member of the sample were tested. Again, using our hypothetical gumball machine, we can draw a sample from this population and then find the sample distribution. If we randomly select 4 gumballs from the population and then note their colors, then we will have the sample distribution which we want. Say, we select at random 1 gumball which is red, 1 which is blue, and 2 which are orange. This would comprise our sample distribution. This sample distribution is graphed below.

As you can see this sample does not describe our population very well. However, there are ways to insure that our sample adequately reflects the population. One of these methods is called the law of large numbers.

Law of large numbers

Definition

Essentially, the law of large numbers states that as the size of the sample increases the sample will more closely approximate the population. Thus bigger samples are more accurate than smaller samples, because the bigger samples more closely represent the population. This is also very appealing to our common sense, because as the sample size gets bigger it includes more of the population, and so it should reflect the true aspects of the population more accurately.

Examples

As a very simple example of this problem let's take an urn filled with gumballs. Let's say that the urn has 50 gumballs in it and that 35 of them are blue gumballs and 15 of them are red gumballs. Thus, the 50 gumballs in the urn are the population and the population consists of 30%

red gumballs and 70% blue gumballs. Let's compare how closely different sample sizes approximate the true proportion of red and blue gumballs in the population. For example let's take 4 samples of size 1, 4 samples of size 4, and 4 samples of size 15. Of course we must be careful to sample randomly from the population. We expect that the 15 unit samples will more accurately reflect the population than the 4 unit samples, and the 4 unit samples will be more accurate than the 1 unit samples.

Samples of Size One

Gumball Color	Sample 1	Sample 2	Sample 3	Sample 4	Average Sample Percent	Population Percent
	red	red	blue	blue		
Red	100%	100%	0%	0%	50%	30%
Blue	0%	0%	100%	100%	50%	70%

Samples of Size Four

Gumball Color	Sample 1	Sample 2	Sample 3	Sample 4	Average Sample Percent	Population Percent
	blue red red red	red red blue blue	blue blue blue blue	blue blue blue red		
Red	75%	50%	0%	25%	38%	30%
Blue	25%	50%	100%	75%	63%	70%

Samples of Size Fifteen

Gumball Color	Sample 1	Sample 2	Sample 3	Sample 4	Average Sample Percent	Population Percent
	blue	blue	blue	blue		
	blue	blue	red	blue		
	blue	red	red	red		
	blue	red	blue	blue		
	blue	red	blue	blue		
	blue	blue	blue	red		
	blue	red	red	blue		
	blue	blue	red	blue		
	red	blue	red	blue		
	red	blue	blue	blue		
	blue	blue	blue	blue		
	red	red	blue	blue		
	red	blue	red	blue		
	red	blue	blue	blue		
	blue	blue	blue	blue		
Red	33%	33%	40%	13%	30%	30%
Blue	67%	67%	60%	87%	70%	70%

As you can see, the 15 unit samples more accurately reflect the true characteristics of the population than do the 4 unit samples, and the 4 unit samples more accurately reflect the population than the 1 unit samples. In fact, this relationship is so strong that it is the only factor which affects how well the sample approximates the population. In other words, the only variable which affects how well the sample describes the population is the size of the sample.

One reason that the law of large numbers is important to learn is that it applies not only to urns and gumballs. To give you an idea of how broad the law of large numbers is, we will give you two situations in which the law of large numbers applies. After you have read the example, we will analyze each situation in term of the law of large numbers. This will begin in a few minutes.

Situation Sports Probabilistic

Jackie was the coach for a large independent swimming club. His usual style was to spend one day with each of his swimmers in intensive practice and then have them spend the next several days working alone on individual drills. Of course the entire team practiced together for 2 hours every morning and every night. Jackie chose which player to work with for each day by randomly choosing a name from his list of athletes. One particular day he drew one of his new swimmers, Laura, from the list. He watched her very closely during the morning team drill and decided that she was a good athlete. But then during swimming practice she became nervous and did very poorly on her first swimming strokes. Jackie concluded that although Laura was a good swimmer, she got nervous under pressure. However, he thought with a little training and good luck her confidence would improve.

What do you think of Jackie's reasoning about Laura's confidence? Explain your answer.

We can analyze Jackie's conclusion using the law of large numbers by thinking of each athlete as having a *population* of good swimming movements. Jackie is interested in those athletes who are very good swimmers, so we can divide the persons swimming movement population into two categories: good swimming movements and others. We can think of the *population distribution* as the percentage or proportion in each category. For many athletes, the population distribution is actually 0% good swimming movements and 100% other: these athletes are simply not very good swimmers.

By watching the group swim, the coach is observing *samples* of each swimmers movements. In the morning team drill the coach is watching two hours of her performance and is taking a very large sample. Later in the day when he watches her first three swim strokes he is taking a much smaller sample. And we know from the law of large numbers that small samples are

very unreliable estimates of the population. When Laura becomes nervous and performs poorly during her 3 swimming trials, it is because she drew some "unlucky gumballs". Also earlier Jackie had concluded that she was a good athlete on the basis of a much larger sample. Therefore Jackie's conclusion about Laura's confidence is unwarranted, especially in light of the earlier evidence about her athletic prowess.

Situation Sports Probabilistic

Margie and Beth have been friends for many years. In the past few months both have taken up running together. Each day the two women flip a coin to decide which route to take and then end their five mile run with a sprint to Beth's house. And, most days, Beth seems to win the sprint. The other day while Margie was shopping she found a new shoe which the salesman claimed would help her to run faster. He said that the new shoe had an air pillow insole and lighter construction which would use less of her energy in running. So she decided to try the new shoe out and, sure enough, she beat Beth in the sprint. Margie figured that the reason that she won the sprint was because of her new shoes with their remarkable new air pillow construction.

Comment on Margie's reasoning and the conclusion she draws. Do you think there are any other reasons why Margie might have won the race?

Margie is trying to draw a conclusion about a certain *population*. We can think of the members of this population as the sprints which Margie and Beth have engaged in before. If we divide the population into *two groups*, days where Margie wins and days where Beth wins, we can think of the *population distribution* as the percentage in each category. Margie has concluded that her new shoes helped her win the sprint against Beth. Her conclusion was based on observing a *sample* of *size=1*, in which the *sample distribution* was 100% for Margie winning the sprint.

Apart from any other considerations, however, the sample distribution for *size=1* is likely to be very different from the population distribution. Even if the *population distribution* is 70 or 80% for Beth winning the sprint, a 1-0 sample split would not be unusual; just as one would not be surprised to draw one red gumball out of an urn with only 20% red gumballs. So Margie's conclusion is quite unwarranted: a larger sample is needed.

Situation Academic Objective

Mark is a teacher at a small mid-western college. During his first year of teaching, he selected a new textbook for his introductory class, An Introduction to Psychology. He likes this new book very much. However, some of his colleagues at a much larger university have been using both his book and a different text, Psychology for the Ages, in their introductory psychology classes. Both colleges have excellent academic reputations and recruit from the same pool of college bound seniors. Mark's colleagues report that Psychology for the Ages is much easier for the students to understand and that the students seem to like the book better.

Should Mark switch to using the new book, Psychology for the Ages, or should he continue to use the one he likes best?

Mark is trying to figure out which book he should use in his courses. To make this decision he has information from both his own classroom work and some of his colleagues at a larger university. According to the law of large numbers, which states that the larger the sample, the better it is in estimating the population, there is substantial evidence that Mark should switch to using the new book, Psychology for the Ages. While Marks' own data indicate that other book, An Introduction to Psychology, is better, that conclusion is based on a smaller sample. So Mark should switch to using the new book, Psychology for the Ages, because a larger group of students have found it be better than the other text.

Situation Academic Objective

Jayne is a professor at a small college in the midwest and she must find a way to finance a field trip that she would like her classes to take. She has narrowed her choice to two fund-raising organizations: The Fund Raisers and Money Makers, INC. Several of her friends have used the Fund Raisers before and they say that this company has done a poor job in helping the students to raise money. Several other friends tell her that Money Makers, INC. is more involved with day-to-day interactions with the students and has helped them to raise several thousand dollars. Jayne also decided to visit the Better Business Bureau and look at each company's report. She found one report from the Fund Raisers which indicated that they had raised nearly ten times as much money as Money Makers, over the same time period.

Analyze the pros and cons for hiring each organization and comment on which firm you think Jayne should hire?

Jayne is trying to draw a conclusion about a certain group of fund raising organizations. If we consider all the firms that Jayne has information then we can think of the reports about each organization as the population. If we divide the population into two groups, organizations who have good reports and organizations who did not have good reports, then we can think of the population distribution as the percentage in each category. Jayne has sampled 1 report from the Better Business Bureaus Files.

Apart from any other considerations, however, the sample distribution for size=1 is likely to be very different from the population distribution. Even if the *population distribution* is 70 or 80% for the Fund Raisers having a better report, a 1-0 sample split would not be unusual; just as one would not be surprised to draw one red gumball out of an urn with only 20% red gumballs. So Jayne's conclusion is quite unwarranted: a larger sample is needed.

APPENDIX D

Stimuli and Read Grade Level Analyses for the Stories Used in Experiment 2.

Table D-1

Stimulus Organization and References for Experiment 2

Identifier	Context	Domain	Phase of Use	Problem Reference
AO1	Academic	Objective	Train	Author
AO2	Academic	Objective	Test	Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). <i>Appendix A, Problem 10</i>
AO4	Academic	Objective	Test	Author
AO5	Academic	Objective	Train	Author
AP1	Academic	Probabilistic	Test	Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). <i>Appendix A, Problem 1</i>
AP2	Academic	Probabilistic	Test	Author
SO1	Sports	Objective	Test	Author
SO2	Sports	Objective	Test	Nisbett, R. E., Krantz, D. H., & Jepson, C., & Kunda, Z. (1983). Page 356—The Championship Selection Problem
SP1	Sports	Probabilistic	Test	Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). <i>Appendix A, Problem 7, rewritten</i>
SP2	Sports	Probabilistic	Test	Author
SP4	Sports	Probabilistic	Train	Author
SP5	Sports	Probabilistic	Train	Author
DO1	Distractor		Test	Kahneman, D., & Tversky, A. (1973).
DO2	Distractor		Test	Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). Page xxx—The Shreeble Problem
DR1	Distractor		Test	Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda Z. (1983). Page 359—The Rookie of the Year
DR3	Distractor		Test	Author

Context: Academic

Source: Author

Domain: Objective

Identification: ao1

Used in: Training

Mark is a teacher at a small mid-western college. During his first year of teaching, he selected a new textbook for his introductory class, An Introduction to Psychology. He likes this new book very much. However, some of his colleagues at a much larger university have been using both his book and a different text, Psychology for the Ages, in their introductory psychology classes. Both colleges have excellent academic reputations and recruit from the same pool of college bound seniors. Mark's colleagues report that Psychology for the Ages is much easier for the students to understand and that the students seem to like the book better.

Should Mark switch to using the new book, Psychology for the Ages, or should he continue to use the one he likes best?

DESCRIPTIVES:

Number of sentences.....	7	Coleman-Liau.....	10.8
Number of words.....	129	Dale-Chall.....	9-10 (25 words not in list)
Number of syllables.....	204	Flesch.....	12.2 (REI is 52.8)
Number of letters.....	619	Flesch-Kincaid.....	10.3
Number of words of 6 or more letters.....	40	Forecast.....	9.0
Number of words of 7 or more letters.....	31	Fry.....	10.0
Number of 1 syllable words.....	89	Gunning-Fog.....	14.2
Number of words of 3 or more syllables.....	22	Holmquist.....	7.2
Percentage of 1 syllable words.....	69.0	Lix.....	9.0 (Lix index is 42)
Percentage of words of 3 or more syllables ...	17.1	Powers.....	6.4
Average number of words per sentence.....	18.4	Raygor.....	10.0
Average number of letters per word.....	4.8	Rix.....	9.0 (Rix index is 4.43)
Average number of syllables per word.....	1.6	McLaughlin-Smog....	13.3

Context:	Academic	Source:	Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). <i>Appendix A, Problem 10</i>
Domain:	Objective	Identification:	ao2
		Used in:	Testing

The Psychology Department of the University of Michigan keeps records on the performance of all of its graduate students and relates this performance score to all kinds of background information about the students. Recently there was a debate among the members of the admissions committee about whether to admit a particular student from Horace Maynard College. The student's scores on the GRE and his GPA were marginal—that is, almost all students actually admitted to the department have scores as high or higher, while most rejected students have lower scores. The student's letters of recommendation were quite good, but none of the writers of the letters were personally known to any of the Michigan faculty.

One member of the admissions committee argued against admission, pointing out that department records show that students who graduate from small non-selective colleges like Maynard perform at a level substantially below the median of all Michigan graduate students. This argument was countered by a committee member who noted that 2 years ago Michigan admitted a student from Maynard who was now among the three highest ranked students in the department.

Comment on the arguments put forward by these two committee members. What are their strengths and weaknesses?

DESCRIPTIVES:

Number of sentences.....	8
Number of words.....	202
Number of syllables.....	341
Number of letters.....	1065
Number of words of 6 or more letters.....	84
Number of words of 7 or more letters.....	72
Number of 1 syllable words.....	115
Number of words of 3 or more syllables.....	41
Percentage of 1 syllable words.....	56.9
Percentage of words of 3 or more syllables..	20.3
Average number of words per sentence.....	25.3
Average number of letters per word.....	5.3
Average number of syllables per word.....	1.7

GRADE LEVELS:

Coleman-Liau.....	14.0
Dale-Chall.....	College (67 words)
Flesch.....	14.9 (REI is 37.3)
Flesch-Kincaid.....	14.2
Forecast.....	11.0
Fry.....	College
Gunning-Fog.....	18.2
Holmquist.....	12.4
Lix.....	College I (Lix index is 60)
Powers.....	7.4
Raygor.....	Scientific/technical
Rix.....	College (Rix index is 9)
McLaughlin-Smog....	16.1

Context: Academic

Source: Author

Domain: Objective

Identification: ao4

Used in: Testing

Greg had taken a course on public speaking in his last semester of graduate school. The professor who taught the course stressed the theory of communication and concentrated on writing skills. After taking the class, Greg felt that he was a much better speaker, was more relaxed, and communicated much more effectively. Before the class he remembered being nervous and irritable before his only presentation. In fact he was so nervous that he practiced his speech for several days and carried a verbatim copy of his notes to the speech. But, everything he did to improve his speech did not help: the faculty ratings of his presentation were still low. However, at several speeches after the class he felt very calm and thought that his presentations went very well. And this time the average faculty ratings of his last six talks were higher than after his first and only speech before he took the class on public speaking. He reasoned that this was because the course prepared him very well for public speaking and had given him the basic skills necessary to communicate.

Comment on the reasoning Greg used in arriving at the conclusion that, "the course prepared him very well for public speaking?" Evaluate arguments for and against his position.

DESCRIPTIVES:

Number of sentences.....	12	Coleman-Liau	10.9
Number of words.....	211	Dale-Chall.....	9-10 (38 words not in list)
Number of syllables.....	325	Flesch.....	9.6 (REI is 62.1)
Number of letters.....	1019	Flesch-Kincaid.....	9.4
Number of words of 6 or more letters	74	Forecast.....	9.6
Number of words of 7 or more letters	51	Fry	9.0
Number of 1 syllable words	139	Gunning-Fog.....	12.3
Number of words of 3 or more syllables	28	Holmquist.....	8.7
Percentage of 1 syllable words.....	65.9	Lix.....	9.0 (Lix index is 41)
Percentage of words of 3 or more syllables ...	13.3	Powers.....	6.2
Average number of words per sentence.....	17.6	Raygor	11.0
Average number of letters per word.....	4.8	Rix	9.0 (Rix index is 4.25)
Average number of syllables per word	1.5	McLaughlin-Smog....	11.9

Context: Academic

Source: Author

Domain: Objective

Identification: ao5

Used in: Training

Jayne is a professor at a small college in the midwest and she must find a way to finance a field trip that she would like her classes to take. She has narrowed her choice to two fund-raising organizations: The Fund Raisers and Money Makers, INC. Several of her friends have used the Fund Raisers before and they say that this company has done a poor job in helping the students to raise money. Several other friends tell her that Money Makers, INC. is more involved with day-to-day interactions with the students and has helped them to raise several thousand dollars. Jayne also decided to visit the Better Business Bureau and look at each company's report. She found that the Fund Raisers had raised nearly ten times as much money as Money Makers, over the same time period. Also the Better Business Bureau had many more complimentary reports about the Fund Raisers than the Money Makers.

Analyze the pros and cons for hiring each organization and comment on which firm you think Jayne should hire?

DESCRIPTIVES:

Number of sentences.....	10	GRADE LEVELS:	
Number of words.....	175	Coleman-Liau	9.8
Number of syllables.....	258	Dale-Chall	9-10 (37 words not in list)
Number of letters	811	Flesch.....	9.6 (REI is 62.2)
Number of words of 6 or more letters	51	Flesch-Kincaid.....	8.6
Number of words of 7 or more letters	34	Forecast.....	9.5
Number of 1 syllable words	116	Fry	9.0
Number of words of 3 or more syllables.....	17	Gunning-Fog.....	10.9
Percentage of 1 syllable words.....	66.3	Holmquist	8.6
Percentage of words of 3 or more syllables	9.7	Lix.....	8.0 (Lix index is 36)
Average number of words per sentence.....	17.5	Powers.....	5.9
Average number of letters per word.....	4.6	Raygor	9.0
Average number of syllables per word	1.5	Rix	8.0 (Rix index is 3.4)
		McLaughlin-Smog.....	10.6

Context:	Academic	Source:	Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). <i>Appendix A, Problem 1</i>
Domain:	Probabilistic	Identification:	ap1
		Used in:	Testing

At Stanbrook University, the Housing Office determines which of the 10,000 students enrolled will be allowed to live on campus the following year. At Stanbrook, the dormitory facilities are excellent, so there is always great demand for on-campus housing. Unfortunately, there are only enough on-campus spaces for 5000 students. The Housing Office determines who will get to live on campus by having a Housing Draw every year: every student picks a number out of a box over a 3-day period. These numbers range from 1 to 10,000. If the number is 5000 or under, the student gets to live on campus. If the number is over 5000, the student will not be able to live on campus.

On the first day of the draw, Joe talks to five people who have picked a number. Of these, four people got low numbers. Because of this, Joe suspects that the numbers in the box were not properly mixed, and that the early numbers are more favorable. He rushes over to the Housing Draw and picks a number. He gets a low number. He later talks to four people who drew their numbers on the second or third day of the draw. Three got high numbers. Joe says to himself, "I'm glad that I picked when I did, because it looks like I was right that the numbers were not properly mixed."

What do you think of Joe's reasoning? Explain.

DESCRIPTIVES:

Number of sentences.....	16	GRADE LEVELS:	
Number of words.....	242	Coleman-Liau	6.8
Number of syllables.....	338	Dale-Chall.....	7-8 (32 words not in list)
Number of letters.....	1011	Flesch.....	7.7 (REI is 73.1)
Number of words of 6 or more letters	65	Flesch-Kincaid.....	6.8
Number of words of 7 or more letters	39	Forecast.....	8.9
Number of 1 syllable words	169	Fry	7.0
Number of words of 3 or more syllables.....	16	Gunning-Fog.....	8.7
Percentage of 1 syllable words.....	69.8	Holmquist.....	7.9
Percentage of words of 3 or more syllables	6.6	Lix.....	6.0 (Lix index is 31)
Average number of words per sentence.....	15.1	Powers.....	5.3
Average number of letters per word.....	4.2	Raygor	7.0
Average number of syllables per word	1.4	Rix	7.0 (Rix index is 2.44)
		McLaughlin-Smog.....	8.8

Context: Academic Source: Author
 Domain: Probabilistic Identification: ap2
 Used in: Testing

Kalman University, a very old institution in the northeast, has had many complaints about the quality of the student canteens. Many of the machines are no longer functioning properly and several have quit working outright. In fact, in a random sample of the fifteen canteens on campus, 10 had at least one machine which was broken. In one particular canteen two soda pop machines stand side-by-side: one of them sells Coke and the other Pepsi. Keith, a senior at Kalman who has visited this canteen many times, has noticed that the Pepsi machine breaks down less often than the Coke machine. Keith notices that Sheila, a new student, is having a hard time deciding which kind of soda pop to buy. He tells her about the finicky Coke machine and suggests that she buy a Pepsi. Sheila tries the Pepsi machine twice and both times it takes her money without returning a Pepsi. Then she tries the Coke machine and it works on the first try. She concludes that Keith must have been mistaken about the two pop machines and the Coke machine is really more reliable than the Pepsi machine.

Comment on Sheila's reasoning. Do you agree? Explain your answer.

DESCRIPTIVES:

Number of sentences.....	14	GRADE LEVELS:	
Number of words.....	201	Coleman-Liau	9.1
Number of syllables.....	303	Dale-Chall.....	9-10 (45 words not in list)
Number of letters	923	Flesch.....	8.9 (REI is 65.3)
Number of words of 6 or more letters	59	Flesch-Kincaid.....	7.8
Number of words of 7 or more letters	45	Forecast.....	10.4
Number of 1 syllable words	122	Fry	8.0
Number of words of 3 or more syllables.....	17	Gunning-Fog.....	9.1
Percentage of 1 syllable words.....	60.7	Holmquist.....	9.3
Percentage of words of 3 or more syllables	8.5	Lix.....	8.0 (Lix index is 36)
Average number of words per sentence.....	14.4	Powers.....	5.8
Average number of letters per word.....	4.6	Raygor	8.0
Average number of syllables per word	1.5	Rix	8.0 (Rix index is 3.21)
		McLaughlin-Smog.....	9.4

Context: Sports Source: Author
 Domain: Objective Identification: so1
 Used in: Testing

June Farrell is an avid fan of her hometown baseball team, the Drunkards, and she has attended almost all of their games for the past several years. While watching one home game with a friend, Helen, she commented on the outstanding quality of one of the outfielders. In particular, she felt that he had great fielding ability and was an excellent batter. Although Helen had not been to many games that year, she thought he was no better than average, and would probably be traded. She said, "Today, he missed two easy fly balls and he grounded out two times at bat."

Comment on Helen and June's reasoning. Explain your answer.

DESCRIPTIVES:		GRADE LEVELS:	
Number of sentences.....	6	Coleman-Liau	8.8
Number of words.....	108	Dale-Chall.....	9-10 (20 words not in list)
Number of syllables.....	157	Flesch.....	9.7 (REI is 61.7)
Number of letters	482	Flesch-Kincaid.....	8.6
Number of words of 6 or more letters	29	Forecast.....	9.3
Number of words of 7 or more letters	23	Fry	9.0
Number of 1 syllable words	73	Gunning-Fog.....	11.6
Number of words of 3 or more syllables.....	12	Holmquist.....	6.6
Percentage of 1 syllable words.....	67.6	Lix.....	8.0 (Lix index is 39)
Percentage of words of 3 or more syllables ...	11.1	Powers.....	5.8
Average number of words per sentence.....	18.0	Raygor	8.0
Average number of letters per word.....	4.5	Rix	9.0 (Rix index is 3.83)
Average number of syllables per word	1.5	McLaughlin-Smog....	11.2

Context: Sports

Source: Nisbett, R. E., Krantz, D. H.,
Jepson, C., & Kunda, Z. (1983).
Page 356, *The Championship
Selection Problem*

Domain: Objective

Identification: so2

Used in: Testing

Two sports fans are arguing over which sport--baseball or football-- has the best (most accurate) play-off system. Charlie says that the Super Bowl is the best way of determining the world champion because, according to him, "the seven games of the world series are all played in the home cities of the two teams, whereas the Super Bowl is usually played in a neutral city. Since you want all factors not related to the game to be equal for a championship, then the Super Bowl is the better way to determine the world championship."

Which procedure do you think is a better way to determine the world champion--World Series or Super Bowl? Why?

DESCRIPTIVES:

Number of sentences.....	5	Coleman-Liau	9.8
Number of words.....	113	Dale-Chall.....	11-12 (28 words not in list)
Number of syllables.....	168	Flesch.....	10.9 (REI is 57)
Number of letters	518	Flesch-Kincaid.....	10.8
Number of words of 6 or more letters	30	Forecast.....	9.2
Number of words of 7 or more letters	21	Fry	10.0
Number of 1 syllable words	77	Gunning-Fog.....	14.0
Number of words of 3 or more syllables.....	14	Holmquist.....	7.8
Percentage of 1 syllable words.....	68.1	Lix.....	9.0 (Lix index is 41)
Percentage of words of 3 or more syllables ...	12.4	Powers.....	6.3
Average number of words per sentence.....	22.6	Raygor	9.0
Average number of letters per word.....	4.6	Rix	9.0 (Rix index is 4.2)
Average number of syllables per word	1.5	McLaughlin-Smog....	12.7

Context: Sports Source: Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). *Appendix A, Problem 7, rewritten*

Domain: Probabilistic Identification: sp1

Used in: Testing

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. He gingerly flips a coin to decide which section of the court he should sit in to watch the game. The player he observes is generally excellent. He repeatedly makes plays worthy of the best professional players. However, in one of the games, with his team behind by 2 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?

DESCRIPTIVES:

Number of sentences.....	11	GRADE LEVELS:	
Number of words.....	213	Coleman-Liau	10.1
Number of syllables.....	315	Dale-Chall	11-12 (46 words not in list)
Number of letters	994	Flesch.....	10.0 (REI is 60.2)
Number of words of 6 or more letters	76	Flesch-Kincaid.....	9.4
Number of words of 7 or more letters	47	Forecast.....	9.3
Number of 1 syllable words	144	Fry	9.0
Number of words of 3 or more syllables.....	25	Gunning-Fog.....	12.5
Percentage of 1 syllable words.....	67.6	Holmquist.....	9.7
Percentage of words of 3 or more syllables	11.7	Lix.....	9.0 (Lix index is 41)
Average number of words per sentence.....	19.4	Powers.....	6.0
Average number of letters per word.....	4.7	Raygor	12.0
Average number of syllables per word	1.5	Rix	9.0 (Rix index is 4.27)
		McLaughlin-Smog.....	11.7

Context: Sports Source: Author
 Domain: Probabilistic Identification: sp2
 Used in: Testing

Jason has been a professional golfer for the past fifteen years. Recently, he and a friend were on the golf course when Jason complained that he had not been hitting the ball as far on his tee shots for the last several months. His friend suggested that he try a new brand of golf ball and he randomly pulled out a new ball from his golf bag. Jason tried the new brand and found that it greatly increased the distance that he got off his tee shots for the day. His friend thought that the increase in distance was due to a new ball construction method which decreased the wind resistance of the ball.

Comment on Jason's friend's reasoning. Do you think that this new ball was the reason Jason's tee shots went farther? Explain your answer.

DESCRIPTIVES:

Number of sentences..... 8
 Number of words..... 137
 Number of syllables..... 176
 Number of letters 594
 Number of words of 6 or more letters 34
 Number of words of 7 or more letters 24
 Number of 1 syllable words 107
 Number of words of 3 or more syllables..... 8
 Percentage of 1 syllable words..... 78.1
 Percentage of words of 3 or more syllables 5.8
 Average number of words per sentence..... 17.1
 Average number of letters per word..... 4.3
 Average number of syllables per word 1.3

GRADE LEVELS:

Coleman-Liau 8.0
 Dale-Chall 9-10 (28 words not in list)
 Flesch..... 7.1 (REI is 79.5)
 Flesch-Kincaid 6.2
 Forecast..... 7.5
 Fry 6.0
 Gunning-Fog 9.2
 Holmquist..... 7.5
 Lix..... 7.0 (Lix index is 34)
 Powers..... 5.0
 Raygor 7.0
 Rix 8.0 (Rix index is 3)
 McLaughlin-Smog..... 8.8

Context: Sports Source: Author
 Domain: Probabilistic Identification: sp4
 Used in: Training

Jackie was the coach for a large independent swimming club. His usual style was to spend one day with each of his swimmers in intensive practice and then have them spend the next several days working alone on individual drills. Of course the entire team practiced together for 2 hours every morning and every night. Bill chose which player to work with for each day by randomly choosing a name from his list of athletes. One particular day he drew one of his new swimmers, Laura, from the list. He watched her very closely during the morning team drill and decided that she was a good athlete. But then during diving practice she became nervous and did very poorly on her first three dives. Jackie concluded that although Laura was a good swimmer, she got nervous under pressure. However, he thought with a little training and good luck her confidence would improve.

What do you think of Jackie's reasoning about Laura's confidence? Explain your answer.

DESCRIPTIVES:		GRADE LEVELS:	
Number of sentences.....	11	Coleman-Liau	9.8
Number of words.....	165	Dale-Chall	9-10 (31 words not in list)
Number of syllables.....	242	Flesch.....	9.1 (REI is 64.7)
Number of letters	774	Flesch-Kincaid.....	7.6
Number of words of 6 or more letters	51	Forecast.....	9.6
Number of words of 7 or more letters	38	Fry	8.0
Number of 1 syllable words	108	Gunning-Fog.....	9.9
Number of words of 3 or more syllables.....	16	Holmquist.....	7.8
Percentage of 1 syllable words.....	65.5	Lix.....	8.0 (Lix index is 38)
Percentage of words of 3 or more syllables	9.7	Powers.....	5.6
Average number of words per sentence.....	15.0	Raygor	9.0
Average number of letters per word.....	4.7	Rix	8.0 (Rix index is 3.45)
Average number of syllables per word	1.5	McLaughlin-Smog.....	10.0

Context: Sports Source: Author
 Domain: Probabilistic Identification: sp5
 Used in: Training

Margie and Beth have been friends for many years. In the past few months both have taken up running together. Each day the two women flip a coin to decide which route to take and then end their five mile run with a sprint to Beth's house. And, most days, Beth seems to win the sprint. The other day while Margie was shopping she found a new shoe which the salesman claimed would help her to run faster. He said that the new shoe had an air pillow insole and lighter construction which would use less of her energy in running. So she decided to try the new shoe out and, sure enough, she beat Beth in the sprint. Margie figured that the reason that she won the sprint was because of her new shoes with their remarkable new air pillow construction.

Comment on Margie's reasoning and the conclusion she draws. Do you think there are any other reasons why Margie might have won the race?

DESCRIPTIVES:		GRADE LEVELS:	
Number of sentences.....	10	Coleman-Liau	7.4
Number of words.....	166	Dale-Chall.....	7-8 (22 words not in list)
Number of syllables.....	211	Flesch.....	7.0 (REI is 80)
Number of letters.....	704	Flesch-Kincaid.....	5.9
Number of words of 6 or more letters.....	37	Forecast.....	7.4
Number of words of 7 or more letters.....	19	Fry.....	6.0
Number of 1 syllable words.....	131	Gunning-Fog.....	8.8
Number of words of 3 or more syllables.....	9	Holmquist.....	6.8
Percentage of 1 syllable words.....	78.9	Lix.....	6.0 (Lix index is 28)
Percentage of words of 3 or more syllables.....	5.4	Powers.....	4.9
Average number of words per sentence.....	16.6	Raygor.....	7.0
Average number of letters per word.....	4.2	Rix.....	6.0 (Rix index is 1.9)
Average number of syllables per word.....	1.3	McLaughlin-Smog.....	8.5

Context: Distractor

Source: Kahneman, D., & Tversky, A.
(1973).

Domain: –

Identification: do1

Used in: Testing

Given that 30% of the men in a sample are engineers and 70% are lawyers, answer the following question.

Jack is a 45 year old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time in his many hobbies which include home carpentry, sailing , and mathematical puzzles.

The probability that Jack is one of the 30 engineers in the sample of 100 is ____%.

Why?

DESCRIPTIVES:

Number of sentences.....	5	GRADE LEVELS:	
Number of words.....	84	Coleman-Liau	8.1
Number of syllables.....	129	Dale-Chall	9-10 (14 words not in list)
Number of letters	367	Flesch.....	9.4 (REI is 62.9)
Number of words of 6 or more letters	26	Flesch-Kincaid.....	9.1
Number of words of 7 or more letters	20	Forecast	9.0
Number of 1 syllable words	58	Fry	9.0
Number of words of 3 or more syllables	12	Gunning-Fog.....	12.4
Percentage of 1 syllable words.....	69.0	Holmquist.....	5.9
Percentage of words of 3 or more syllables ...	14.3	Lix.....	9.0 (Lix index is 40)
Average number of words per sentence	16.8	Powers.....	6.1
Average number of letters per word.....	4.4	Raygor	9.0
Average number of syllables per word	1.5	Rix	9.0 (Rix index is 4)
		McLaughlin-Smog....	12.0

Context: Source: Nisbett, R. E., Krantz, D. H., 118
Jepson, C. Kunda, Z. (1983).
Page xxx—*The Shreeble Problem*

Domain: Identification: do2

Used in: Testing

Imagine that you are an explorer who has landed on a little known island in the Southeastern Pacific. You encounter several new animals, plants, and objects. You observe the properties of the animals you meet and you need to make guesses about how common these properties would be in other animals. Suppose you encounter a new bird, the shreeble. It is blue in color.

How many shreebles do you think are this color and why?

DESCRIPTIVES:

Number of sentences.....	6	GRADE LEVELS:	
Number of words.....	75	Coleman-Liau	8.5
Number of syllables.....	114	Dale-Chall.....	7-8 (13 words not in list)
Number of letters	340	Flesch.....	8.6 (REI is 67.2)
Number of words of 6 or more letters	23	Flesch-Kincaid.....	7.2
Number of words of 7 or more letters	18	Forecast.....	9.9
Number of 1 syllable words	48	Fry	8.0
Number of words of 3 or more syllables	12	Gunning-Fog.....	11.4
Percentage of 1 syllable words.....	64.0	Holmquist.....	5.6
Percentage of words of 3 or more syllables ...	16.0	Lix.....	8.0 (Lix index is 36)
Average number of words per sentence	12.5	Powers.....	5.7
Average number of letters per word.....	4.5	Raygor	8.0
Average number of syllables per word	1.5	Rix	8.0 (Rix index is 3)
		McLaughlin-Smog....	11.2

Context: Distractor

Source: Nisbett, R. E., Krantz, D. H.,
Jepson, C. Kunda, Z. (1983).
Page 359—The Rookie of the
Year

Domain: —

Identification: dr1

Used in: Testing

In general the major league baseball player who wins Rookie of the Year does not perform as well in his second year. This is clear in major league baseball in the past ten years. In the American League, 8 rookies of the year have done worse in their second year; only 2 have done better. In the National League, the rookie of the year has done worse the second year 9 times out of 10.

Why do you suppose the Rookie of the Year tends not to do as well his second year?

DESCRIPTIVES:

Number of sentences.....	5
Number of words.....	93
Number of syllables.....	122
Number of letters.....	366
Number of words of 6 or more letters.....	21
Number of words of 7 or more letters.....	8
Number of 1 syllable words.....	68
Number of words of 3 or more syllables.....	3
Percentage of 1 syllable words.....	73.1
Percentage of words of 3 or more syllables.....	3.2
Average number of words per sentence.....	18.6
Average number of letters per word.....	3.9
Average number of syllables per word.....	1.3

GRADE LEVELS:

Coleman-Liau.....	5.7
Dale-Chall.....	7-8 (11 words not in list)
Flesch.....	7.2 (REI is 78)
Flesch-Kincaid.....	7.1
Forecast.....	8.4
Fry.....	7.0
Gunning-Fog.....	8.7
Holmquist.....	5.6
Lix.....	5.0 (Lix index is 27)
Powers.....	5.2
Raygor.....	7.0
Rix.....	5.0 (Rix index is 1.6)
McLaughlin-Smog.....	7.6

Context: Distractor

Source: Author

Domain: –

Identification: dr3

Used in: Testing

Joe is the retiring Dean of a large social sciences school. In his exit interview he reports the following of his experiences at the university: "One of the school's biggest problems is the turnover rate among new professors. Each year we hire between 5 and 10 new educators. A few of these men and women are very ambitious—they seem eager and enthusiastic about their new positions. As an example, one of them wrote a new textbook during her first year, while another played a major part in overhauling the graduate and undergraduate curricula. But, I usually find that after several years, these professors are not much more energetic and eager than the others. I think that they put so much effort in at the beginning of their careers that they get burned out very quickly. Usually they continue to do research and contribute to the program at a normal rate, but nothing like what their first year pace would have suggested."

Comment on Joe's reasoning. Do you think it is likely that he correctly identifies the professors who are eager and ambitious? Do you agree it is likely that most of the ones who are very eager at first later become burned out?

DESCRIPTIVES:

Number of sentences.....	12	GRADE LEVELS:	
Number of words.....	204	Coleman-Liau	9.4
Number of syllables.....	313	Dale-Chall	9-10 (40 words not in list)
Number of letters	933	Flesch.....	9.5 (REI is 62.7)
Number of words of 6 or more letters	56	Flesch-Kincaid.....	9.1
Number of words of 7 or more letters	44	Forecast	9.3
Number of 1 syllable words	138	Fry	9.0
Number of words of 3 or more syllables.....	31	Gunning-Fog.....	12.9
Percentage of 1 syllable words.....	67.6	Holmquist	8.9
Percentage of words of 3 or more syllables ...	15.2	Lix.....	8.0 (Lix index is 38)
Average number of words per sentence.....	17.0	Powers.....	6.1
Average number of letters per word.....	4.6	Raygor	8.0
Average number of syllables per word	1.5	Rix	8.0 (Rix index is 3.67)
		McLaughlin-Smog....	12.3

APPENDIX E

Phone Consent Form and Survey Questions Used in Experiment 2.

Telephone Consent Form and Survey Questions

[THIS WILL BE READ TO ALL SUBJECTS WHO ARE CONTACTED BY PHONE.]

"Hello my name is _____ from Rice University. I am conducting a survey about some current campus issues at Rice University."

I need to speak with _____. Is _____ available?

[IF TIME IS NOT CONVENIENT OR PERSON IS NOT AVAILABLE, ASK:]

"Is there another time I can call back?"

[IF PERSON AGREES TO CONTINUE, ASK:]

"The following information will be used only for the purposes of this survey and will not be released to any individual or group in such a way that you can be identified. We are interested in your opinions of several different campus issues. At the conclusion of the survey the information which links your name to your coded responses will be destroyed. Also your responses will be kept in the strictest of confidence. Will you participate in this study?"

[IF PERSON DECLINES SAY:]

"Thank you for your time."

[IF PERSON AGREES TO CONTINUE, ASK:]

"All responses from this survey will be kept in complete confidence and if you allow me to tape record your responses the tape will be destroyed at the end of the experiment. Do you mind if I tape record your responses?"

[IF OK TO TAPE THEN TAPE. IF NOT THEN WRITE ANSWERS TO THE AO AP SO AND SP QUESTIONS ONLY.]

First, I'd like to ask you a few questions for background information only.

1. In which of the following age groups would you place yourself?

[READ LIST, RECORD.]

- A. Under 18
- B. 18-20
- C. 21-23
- D. 24-26
- E. 26-28
- F. Above 28

2. What year of school are you currently in?

[READ LIST, RECORD.]

- A. Freshman
- B. Sophomore
- C. Junior
- D. Senior
- E. Graduate
- F. Other

3. In what ethnic group do you place yourself?

[READ LIST, RECORD.]

- A. White
- B. Black
- C. Hispanic
- D. Asian
- E. Other

I'd like to get your opinion on the following issues.

1. Do you feel the tuition hike will adversely affect the majority of students at Rice University?
2. If the tuition hike is not implemented do you feel the quality of education at Rice University will suffer?
3. The Fondren Library has recently completed an expensive renovation. Has this renovation made the library easier for you to use?
4. The issue of sex discrimination among the faculty, staff, and students of Rice University has prompted serious questions concerning the hiring, promotion, and grading of women at Rice. Do you feel that there are major inequalities in the treatment of men and women at all levels (undergraduate, graduate, faculty, staff, and administrative) of Rice University? Why?
5. Recently, the issue of campus crime has been increasingly discussed among the students. Do you feel the administration should change its current policy toward campus crime, either through increasing the number of police officers or through more diligent monitoring of university entrances?
6. Do you feel that additional campus lighting would reduce the number of crimes committed on campus?
7. Since the new food service system has been in place, have you noticed any improvement in the quality of the food served? Do you have any suggestions for increasing the quality of food service?
8. As far as the issue of student activities is concerned do you feel that the student body of Rice University is apathetic?
9. Since the addition of the MacIntosh II computers to Mudd Laboratory, have you been using the university mainframe computer more or less or has there been no change? Do you feel the money spent on these computers has been worthwhile?
10. In your undergraduate career at Rice University have you found the administrative staff to be helpful and friendly or not? Why?

- AO. Do you feel that your own first impression of a course or the class ratings are more accurate in determining which courses are worth taking? Why or why not?
- AP. Given that the questions on the SAT are randomly arranged, do you think it is fair for students to be admitted to their respective universities on the basis of this single test? Why or why not?
- SO. How do you feel the performance of the Rice University football team will compare with past years, even though the order of the teams they play against is essentially random? Why or why not?
- SP. Do you feel the Rice university baseball team is in for a major turnaround based on their brilliant performance against Texas A&M last season? Why or why not?

APPENDIX F

ANOVA Tables, Means, and Simple Main Effects for the Test Problem and Phone Survey

Analyses of Experiment 2.

Table F-1

Demographics - Descriptive Statistics

Variable	Number	Percentage	Variable	Number	Percentage
Gender			Age		
Female	33	69	17	2	4
Male	15	31	18	20	42
Race			19	10	21
White	33	69	20	8	17
Black	4	8	21	5	10
Hispanic	4	8	22	1	2
Asian	6	13	30	1	2
Indian	1	2	37	1	2
Major			Statistical Courses		
Business	8	23	None	18	50
Engineering	4	12	1 Course	8	22
Humanities	5	15	2 Courses	9	25
Natural Science	8	24	3 Courses	1	3
Social Science	9	26	Missing	12	
Missing	14		Training Condition		
Members in the Training Group			SPT	12	25
2 group members	3	23	AOT	14	29
3 group members	2	15	CT	12	25
4 group members	4	31	NT	10	21
5 group members	4	31			

Table F-2

Test Problem Analysis: Analysis of Variance

Variable		Levels					
Dependent Measure:	Statistical Score	1	2	3			
Independent Measures:	Training Condition	<u>SPT</u>	<u>AOT</u>	<u>CT</u>	<u>NT</u>		
	Mean	2.05	1.86	1.87	1.50		
	SD	0.88	0.88	0.91	0.70		
	Test Phase	<u>Test 1</u>	<u>Test 2</u>				
	Mean	1.92	1.76				
	SD	0.88	0.87				
	Test Problem	<u>AQ</u>	<u>AP</u>	<u>SO</u>	<u>SP</u>	<u>DQ</u>	<u>DR</u>
	Mean	1.81	2.41	2.36	1.97	1.23	1.24
	SD	0.84	0.75	0.78	0.88	0.57	0.54

Source	df	Type III Sums of Squares	Mean Squares	F Value	Probability	Huyn-Feldt Adjusted Probability
Training Condition	3	20.398	6.799	3.75	0.0175	
Error	44	79.761	1.813			
Test Phase	1	3.007	3.007	6.01	0.0183	
Test Phase * Training Condition	3	2.143	0.714	1.43	0.2474	
Error	44	22.016	0.500			
Test Problem	5	126.179	25.236	82.85	0.0001	0.0001
Test Problem * Training Condition	15	8.201	0.547	1.79	0.0366	0.0366
Error	220	67.014	0.305			
Test Phase * Test Problem	5	1.762	0.352	0.77	0.570	ns
Test Phase * Test Problem * Training Condition	15	4.048	0.269	0.59	0.879	ns
Error	220	100.292	0.456			

Table F-3

Analysis of the Simple Effects for the Test Problem by Training Condition Interaction

Effect	Mean Square Effect	Degrees of Freedom Effect	Degrees of Freedom Error	Mean Square Error	F Value	p Value
Training Condition @ Test Problem AO	3.078	3	44	0.760	4.052	0.013
Training Condition @ Test Problem AP	1.717	3	44	0.725	2.366	0.084
Training Condition @ Test Problem DO	0.203	3	44	0.281	0.722	0.544
Training Condition @ Test Problem DR	0.586	3	44	0.280	2.093	0.115
Training Condition @ Test Problem SO	0.940	3	44	0.596	1.579	0.208
Training Condition @ Test Problem SP	3.831	3	44	0.694	5.522	0.003
Test Problem @ Training Condition SPT	4.345	5	220	0.305	14.265	0.001
Test Problem @ Training Condition AOT	3.560	5	220	0.305	11.687	0.001
Test Problem @ Training Condition CT	3.783	5	220	0.305	12.420	0.001
Test Problem @ Training Condition NT	2.060	5	220	0.305	6.763	0.001

Table F-4

Analysis of the Simple Effects for the Test of Transfer of Training

Effect	Mean Square Effect	Degrees of Freedom Effect	Degrees of Freedom Error	Mean Square Error	F Value	p Value
Training Condition @ Test Problem AO	0.096	2	35	0.891	0.108	0.898
Training Condition @ Test Problem AP	0.778	2	35	0.785	0.991	0.381
Training Condition @ Test Problem SO	0.017	2	35	0.622	0.027	0.973
Training Condition @ Test Problem SP	2.637	2	35	0.729	3.615	0.037
Test Problem @ Training Condition SPT	1.047	3	105	.301	3.473	0.019
Test Problem @ Training Condition AOT	1.548	3	105	0.301	5.134	0.002
Test Problem @ Training Condition CT	0.630	3	105	0.301	2.091	0.106

Table F-5

Test Problem Analysis: Sample Size, Means, and Standard Deviations for Test 1

	Academic Objective	Academic Probabilistic	Sports Objective	Sports Probabilistic	Distractor Other	Distractor Regression	Total
Academic Objective							
N	14	14	14	14	14	14	14
Mean	2.14	2.36	2.57	2.07	1.43	1.43	2.00
SD	0.86	0.84	0.65	0.92	0.76	0.76	0.89
Sports Probabilistic							
N	12	12	12	12	12	12	12
Mean	2.08	3.00	2.58	2.58	1.25	1.33	2.14
SD	0.67	0.00	0.67	0.67	0.62	0.65	0.88
Combined							
N	12	12	12	12	12	12	12
Mean	2.17	2.58	2.42	2.25	1.08	1.33	1.97
SD	0.94	0.51	0.90	0.75	0.29	0.49	0.87
None							
N	10	10	10	10	10	10	10
Mean	1.20	2.00	2.10	1.50	1.00	1.00	1.47
SD	0.42	0.82	0.74	0.71	0.00	0.00	0.70
N	48	48	48	48	48	48	
Mean	1.91	2.51	2.43	2.13	1.21	1.30	
SD	0.83	0.72	0.75	0.84	0.55	0.59	

Table F-6

Test Problem Analysis: Sample Size, Means, and Standard Deviations for Test 2

	Academic Objective	Academic Probabilistic	Sports Objective	Sports Probabilistic	Distractor Other	Distractor Regression	Total
Academic Objective							
N	14	14	14	14	14	14	14
Mean	1.64	2.43	2.36	1.57	1.21	1.14	1.73
SD	0.74	0.76	0.93	0.76	0.58	0.36	0.86
Sports Probabilistic							
N	12	12	12	12	12	12	12
Mean	1.92	2.42	2.33	2.33	1.25	1.50	1.96
SD	1.00	0.79	0.78	0.78	0.62	0.80	0.90
Combined							
N	12	12	12	12	12	12	12
Mean	1.83	2.25	2.42	1.83	1.33	1.00	1.78
SD	0.94	0.97	0.79	1.03	0.78	0.00	0.94
None							
N	10	10	10	10	10	10	10
Mean	1.30	2.10	2.00	1.50	1.20	1.10	1.53
SD	0.48	0.57	0.82	0.85	0.42	0.32	0.70
N	48.00	48.00	48.00	48.00	48.00	48.00	
Mean	1.68	2.32	2.30	1.83	1.26	1.19	
SD	0.84	0.79	0.83	0.90	0.61	0.50	

Table F-7

Phone Survey: Analysis of Variance

<u>Variable</u>		<u>Levels</u>			
Dependent Measure: Statistical Score		1	2	3	
Independent Measures: Training Condition		<u>SPT</u>	<u>AQT</u>	<u>CI</u>	<u>NI</u>
Mean		2.01	1.85	1.92	1.42
SD		0.91	0.93	0.93	0.65
Test Phase		<u>Test 1</u>	<u>Phone</u>	<u>Test 2</u>	
Mean		2.27	1.05	2.15	
SD		0.84	0.21	0.88	
Test Problem		<u>AQ</u>	<u>AP</u>	<u>SO</u>	<u>SP</u>
Mean		1.62	2.04	1.92	1.72
SD		0.83	0.92	0.92	0.88

Source	df	Type III Sums of Squares	Mean Squares	F Value	Probability	Huyn-Feldt Adjusted Probability
Training Condition	3	13.360	4.453	2.84	0.0615	
Error	22	34.527	1.569			
Test Phase	2	86.984	43.492	71.38	0.0001	0.0001
Test Phase * Training Condition	6	8.812	1.468	2.41	0.0422	0.0422
Error	44	26.809	0.609			
Test Problem	3	8.494	2.831	15.41	0.0001	0.0001
Test Problem * Training Condition	9	2.668	0.296	1.61	0.1295	ns
Error	66	12.124	0.184			
Test Phase * Test Problem	6	4.351	0.725	2.44	0.0286	0.0438
Test Phase * Test Problem * Training Condition	18	6.433	0.357	1.20	0.2679	ns
Error	132	39.227	0.297			

Table F-8

Analysis of the Simple Effects for the Test Phase by Training Condition Interaction

Effect	Mean Square Effect	Degrees of Freedom Effect	Degrees of Freedom Error	Mean Square Error	F Value	p Value
Training Condition @ Test Phase 1	7.637	3	22	0.946	7.790	0.001
Training Condition @ Test Telephone Survey	0.026	3	22	0.065	0.390	0.762
Training Condition @ Test Phase 2	1.344	3	22	1.777	0.757	0.530
Test Phase @ Training Condition SP	4.514	2	44	0.609	7.408	0.002
Test Phase @ Training Condition AO	3.831	2	44	0.609	6.287	0.004
Test Phase @ Training Condition Combined	3.771	2	44	0.609	6.189	0.004
Test Phase @ Training Condition None	0.804	2	44	0.609	1.320	0.278

Table F-9

Analysis of the Simple Effects for the Test Phase by Test Problem Interaction

Effect	Mean Square Effect	Degrees of Freedom Effect	Degrees of Freedom Error	Mean Square Error	F Value	p Value
Test Problems @ Test Phase 1	1.588	3	66	0.405	3.924	0.012
Test Problems @ Telephone Survey	0.047	3	66	0.039	1.197	0.318
Test Problems @ Test Phase 2	2.647	3	66	0.334	7.915	0.001
Test Phase @ Test Problem AO	5.975	2	44	0.424	14.090	0.001
Test Phase @ Test Problem AP	15.766	2	44	0.252	62.504	0.001
Test Phase @ Test problem SO	15.428	2	44	0.390	39.543	0.001
Test Phase @ Test Problem SP	8.498	2	44	0.434	19.567	0.001

Table F-10

Phone Survey Analysis: Sample Size, Means, and Standard Deviations for Test 1

	Academic Objective	Academic Probabilistic	Sports Objective	Sports Probabilistic	Total
Academic Objective					
N	8	8	8	8	
Mean	2.25	2.38	2.63	2.13	2.34
SD	0.89	0.92	0.74	0.99	0.87
Sports Probabilistic					
N	6	6	6	6	
Mean	2.33	3.00	2.83	2.67	2.71
SD	0.52	0.00	0.41	0.82	0.55
Combined					
N	7	7	7	7	
Mean	2.14	2.86	2.43	2.14	2.39
SD	0.90	0.38	0.98	0.69	0.79
None					
N	5	5	5	5	
Mean	1.20	1.60	1.80	1.20	1.45
SD	0.45	0.89	0.45	0.45	0.60
N	26	26	26	26	
Mean	2.04	2.50	2.46	2.08	2.27
SD	0.82	0.81	0.76	0.89	

Table F-11

Phone Survey Analysis: Sample Size, Means, and Standard Deviations for Telephone Survey

	Academic Objective	Academic Probabilistic	Sports Objective	Sports Probabilistic	Total
Academic Objective					
N	8	8	8	8	
Mean	1.00	1.13	1.00	1.13	1.06
SD	0.00	0.35	0.00	0.35	0.25
Sports Probabilistic					
N	6	6	6	6	
Mean	1.17	1.00	1.00	1.00	1.04
SD	0.41	0.00	0.00	0.00	0.20
Combined					
N	7	7	7	7	
Mean	1.00	1.29	1.00	1.00	1.07
SD	0.00	0.49	0.00	0.00	0.26
None					
N	5	5	5	5	
Mean	1.00	1.00	1.00	1.00	1.00
SD	0.00	0.00	0.00	0.00	0.00
N	26	26	26	26	
Mean	1.04	1.12	1.00	1.04	1.05
SD	0.20	0.33	0.00	0.20	

Table F-12

Phone Survey Analysis: Sample Size, Means, and Standard Deviations for Test 2

	Academic Objective	Academic Probabilistic	Sports Objective	Sports Probabilistic	Total
Academic Objective					
N	8	8	8	8	
Mean	1.75	2.63	2.50	1.75	2.16
SD	0.89	0.74	0.93	0.89	0.92
Sports Probabilistic					
N	6	6	6	6	
Mean	1.83	2.83	1.83	2.67	2.29
SD	0.98	0.41	0.75	0.52	0.81
Combined					
N	7	7	7	7	
Mean	2.14	2.29	2.57	2.14	2.29
SD	1.07	0.95	0.79	1.07	0.94
None					
N	5	5	5	5	
Mean	1.20	2.20	2.20	1.60	1.80
SD	0.45	0.45	0.84	0.89	0.79
N	26	26	26	26	
Mean	1.77	2.50	2.31	2.04	2.15
SD	0.91	0.71	0.84	0.92	