5 The Effects of Statistical Training on Thinking about Everyday Problems

Geoffrey T. Fong
Northwestern University

Richard E. Nisbett
The University of Michigan

David H. Krantz
Columbia University

Do people solve inferential problems in everyday life by using abstract inferential rules or do they use only rules specific to the problem domain? The view that people possess abstract inferential rules and use them to solve even the most mundane problems can be traced back to Aristotle. In modern psychology, this view is associated with the theories of Piaget and Simon. They hold that, over the course of cognitive development, people acquire general and abstract rules and schemas for solving problems. For example, people acquire rules that correspond to the laws of formal logic and the formal rules of probability theory. Problems are solved by decomposing their features and relations into elements that are coded in such a way that they can make contact with these abstract rules.

This formalist view has been buffeted by findings showing that people violate the laws of formal logic and the rules of statistics. People make serious logical errors when reasoning about arbitrary symbols and relations (for a review, see Evans, 1982). The best known line of research is that initiated by Wason (1966) on his selection task. In that task, subjects are told that they will be shown cards having a letter on the front and a number on the back. They are then presented with cards having an A, a B, a 4, and a 7 and asked which they would have to turn over in order to verify the rule, "If a card has an A on one side, then it has a 4 on the other." This research showed that people do not reason in accordance with the simple laws of conditional logic, which would require turning over the A and the 7. Subsequent work showed that people do reason in accordance with the conditional for certain concrete and familiar problems. For example, when people are given envelopes and asked to verify the rule, "If the letter is sealed, then it has a 50-lire stamp on it," they have no trouble with the problem (Johnson-Laird, Legrenzi, & Sonino-Legrenzi, 1972). Many investigators have concluded from results of the latter sort that people do not use abstract rules of logic when solving concrete problems. Instead, people use only domain-specific rules (e.g., D'Andrade, 1982; Golding, 1981; Griggs & Cox, 1982; Johnson-Laird et al., 1972; Manktelow & Evans, 1979; Reich & Ruth, 1982). If people solve a problem correctly, it is because they are sufficiently familiar with the content domain to have induced a rule that allows them to solve problems in that domain.

Research on inductive reasoning has followed a similar history. Kahneman and Tversky (e.g., 1971, 1973; Tversky & Kahneman, 1974) demonstrated that people fall prey to a multitude of failures to employ statistical rules when reasoning about everyday life problems. In particular, people often fail to reason in accordance with the law of large numbers, the regression principle, or the base rate principle. (For reviews see Einhorn & Hogarth, 1981; Hogarth, 1980; Kahneman, Slovic, & Tversky, 1982; Nisbett & Ross, 1980).

We and our colleagues, however, have shown that people do use statistical concepts in solving particular kinds of problems in particular domains (Jepson, Krantz, & Nisbett, 1983; Nisbett, Krantz, Jepson, & Fong, 1982; Nisbett, Krantz, Jepson, & Kunda, 1983). For example, Jepson et al. (1983) presented subjects with a variety of problems drawn from three very broad domains. All of the problems dealt with events that are variable and, as such, can be analyzed in terms of statistical concepts such as sample size. One domain examined by Jepson et al. consisted of problems for which the random nature of the sample is obvious. In one problem, for example, the protagonist has to judge characteristics of a lottery. As expected, the great majority of the answers for these "probabilistic" problems were statistical answers, that is, they incorporated intuitive notions of the law of large numbers or the regression principle in their answer. At the other extreme, a different group of problems dealt with subjective judgments about the properties of some object or person. In one of these problems, for example, the protagonist has to decide which of two college courses he should take, either on the basis of one visit to each class or on the basis of the evaluations of students who took the courses the previous term. Statistical responses were relatively rare for these "subjective" problems, constituting only about a quarter of the total. In between these extremes, there were a number of problems that, while not containing broad hints as to the random nature of the events in question, dealt with events that are of a sufficiently objective nature that it is relatively easy to recognize that they are characterized by a degree of random variation. These problems dealt primarily with athletic events and academic achievements. For these "objective" problems, slightly more than half of the answers were statistical in nature.

Nisbett et al. (1983) interpreted these and similar results as reflecting the fact that people possess intuitive but abstract versions of statistical rules. They called these intuitive rules "statistical heuristics," and argued that people call on such heuristics to the degree that (a) problem features are readily coded in terms of statistical rules, that is, when the sample space and sampling process are clear, and when the events can be coded in common units (as is the case for athletic events and academic achievements, for example); (b) the presence of chance factors or random variation is signaled by the nature of the events or by other cues in the problem; and (c) the culture recognizes the events in question as being associated with random variation (for example, gambling games) and thus prescribes that an adequate explanation of such events should make reference to statistical principles.

This account presumes that statistical heuristics are abstract. It explains people's frequent failures to use abstract rules as being the result of difficulty in coding problem elements in terms that trigger the rules or as the result of the presence of competing heuristics. But the evidence to date does not rule out the view that statistical heuristics are not abstract at all, but rather are local, domain-bound rules that happen to overlap with formal statistical rules. These rules are better developed in some domains than in others, and it is for this reason that people are much more likely to give statistical answers for some problems than others.

If statistical heuristics are abstract, then it should be possible to improve people's statistical reasoning about everyday events by formal instruction in the rule system, without reference to any domain of everyday events. Such abstract instructional methods should help people apply the rules over a broad range of problem content. On the other hand, if such formal instruction fails to help people to solve concrete problems, despite the fact that people can be shown to have learned a substantive amount about the formal properties of the rules, this would be discouraging to the formal view. It would also be discouraging to the formal view if it were to turn out that abstract instruction affects only people's solution of probabilistic problems, where the relevance of statistical rules is obvious, and where competing rules have relatively little strength.

In order to test the view that formal training per se results in an increase in people's use of statistical principles across a variety of domains, we trained subjects, in brief but intensive laboratory sessions, on the concepts associated with the law of large numbers. We then presented them with a number of problems in each of three broad domains, dealing, respectively, with events generally construed as probabilistic, with objectively measurable events, and with events that are measurable only by subjective judgments.

We also tested the formal view in another way. Some subjects were not given formal instruction, but instead were shown how to apply the law of large numbers for three concrete example problems, all of which dealt with objectively measurable events. If subjects are capable of inducing generalized rules of some degree of abstraction from such training, then they might be expected to reason more statistically about problems in the other domains as well, even though they have not been presented with examples in those domains. Whereas the empirical view suggests that statistical training will be domain specific, with training in one domain failing to generalize to other domains, the formalist view predicts that statistical training in one domain should generalize readily to other domains.

All of the problems presented to subjects concerned everyday life events and were of a type that, in previous work, we have found at least some subjects answer in a statistical fashion. All questions were open ended, and we coded the written answers according to a system that distinguished among varying degrees of statistical thinking. This procedure provided us with a great deal of information about how people reason about events in everyday life and allowed us to determine whether training can enhance not only the likelihood of employing statistical concepts, but also the likelihood that those concepts will be employed properly.

EXPERIMENT 1

Testing Method

Subjects' intuitive use of statistical reasoning was tested by examining their answers to 15 problems to which the law of large numbers could be applied and 3 for which the law of large numbers was not relevant. In this section we describe the instructions that introduced the test problems, the design of the 18 problems, and the system of coding the open-ended answers. The actual text of the problems is given in Appendix A.

Instructions

The instructions for the control subjects read as follows:

> We are interested in studying how people go about explaining and predicting events under conditions of very limited information about the events. It seems to us to be important to study how people explain and predict under these conditions because they occur very frequently 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.
>
> On the pages that follow, there are a number of problems that we would like you to consider. As you will see, they represent a wide range of real-life situations. We would like you to think carefully about each problem, and then write down answers that are sensible to you.

For groups that received training, the first paragraph of the above instructions was presented as part of the introduction to the training materials. After the training, the test booklet was introduced by the second paragraph, which ended with the sentence, "In many of the problems, you may find that the Law of Large Numbers is helpful."

Problem Types and Problem Structure

The 18 problems were divided into three major types as follows:

Type 1. Probabilistic. In these six problems, subjects had to draw conclusions about the characteristics of a population from sample data generated in a way that clearly incorporated random variation. Randomness was made clear in various ways: by the explicitly stated variation in sample outcomes (for example, the number of perfect welds out of 900 made by a welding machine ranged from 680 to 740), by including in the problem a random generating device (for example, shaking a jar of pennies before drawing out a sample), or by simply stating that a sample was "random."

Type 2. Objective. In these six problems, subjects had to draw conclusions about characteristics of a population on the basis of "objective" sample data but with no explicit cue about randomness of the data. One problem, for example, asked subjects to decide which of two makes of car was more likely to be free of troublesome repairs, on the basis of various facts about the repair records. Other problems dealt with the outcomes of athletic events and with academic accomplishments.

Type 3. Subjective. In these six problems, subjects had to draw conclusions about subjective characteristics of a population from "subjective" sample data. In one problem, for example, a high school senior had to choose between two colleges. The underlying subjective characteristic in this problem was liking for the two schools and the data consisted of his own and his friends' reactions to the schools.

In order to systematize the kinds of problems we presented to subjects across the three domains, we selected six different underlying problem structures and for each structure we wrote one problem of each of the above three types. The structures varied in types of samples drawn, type of decision required, and type of competing information.

Structure 1 problems required subjects to draw conclusions about a population from a single small sample. Structure 2 problems pitted a small sample against a large sample. Structure 3 problems required subjects to explain why an outcome selected because of its extreme deviation was not maintained in a subsequent sample (i.e., regression). Structure 4 problems were similar to those in Structure 2, except that the large sample was drawn from a population that was related to, although not identical to, the target population. Structure 5 problems pitted a large sample against a plausible theory that was not founded on data. Structure 6 (false alarm) problems involved conclusions drawn from a sample that was large, but also highly biased. As such, criticism or arguments in these problems should be based on the sample *bias,* but not on sample size. We included these problems to determine whether subjects who received training on the law of large numbers would then proceed to invoke it indiscriminately, or if they would apply it only to the problems of Structures 1–5, for which it was genuinely relevant.

In short, the 18 test problems followed a 3×6 design, with problem type crossed with problem structure. The order of the 18 test problems was randomized for each subject, with the constraint that no 2 problems with the same structure appeared successively.

Coding System

To study the use of statistical reasoning, a simple 3-point coding system was developed for the 15 problems for which the law of large numbers was applicable (Structures 1–5). To illustrate this coding system, we present examples of responses to the "slot machine problem," the probabilistic version of Structure 2 (small sample vs large sample). The protagonist of the story, Keith, was in a Nevada gas station where he played two slot machines for a couple of minutes each day. He lost money on the left slot machine and won money on the right slot machine. Keith's result, however, ran counter to the judgment of an old man sitting in the gas station, who said to Keith, "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." Keith's conclusion after playing the slot machines was that the old man was wrong about the chances of winning on the two slot machines. Subjects were asked to comment on Keith's conclusion. Every response to the test problems was classified into one of three categories:

1 = an entirely deterministic response, that is, one in which the subject made no use of statistical concepts. In responses of this type, there was no mention of sample size, randomness, or variance. The following was coded as a deterministic response to the slot machine problem: "Keith's reasoning was poor, provided the information given by the man was accurate. The man, however, may have been deceiving Keith."

2 = a poor statistical response. Responses given this score contained some mention of statistical concepts, but were incomplete or incorrect. These responses contained one or more of the following characteristics: (1) the subject used both deterministic and statistical reasoning, but the deterministic reasoning was judged by the coder to have been preferred by the subject; (2) the subject used incorrect statistical reasoning, such as the gambler's fallacy; (3) the subject mentioned luck or chance or the law of large numbers but was not explicit about how the statistical concept was relevant. The following is an example of a poor statistical response to the slot machine problem:

> I think that Keith's conclusion is wrong because the old man had better luck on the left one, so he thought it was better. Keith had better luck on the right one so he thought it was better. I don't think you could have a better chance on either one.

3 = a good statistical response. Responses given this score made correct use of a statistical concept. Some form of the law of large numbers was used, and the sampling elements were correctly identified. If the subject used both deterministic and statistical reasoning, the statistical reasoning was judged by the coder to have been preferred by the subject. In general, the subject was judged to have clearly demonstrated how the law of large numbers could be applied to the problem. The following was coded as a good statistical response to the slot machine problem:

> Keith's conclusion is weak. He is wrong in making the assumptions against the old man. Keith is judging the machines on only a handful of trials and not with the sample number the old man has developed over the years. Therefore, Keith's margin of error is much more great than the old man's.

The coding system thus distinguished each response on the basis of whether or not a statistical concept had been used and, within the class of statistical responses, whether or not it was a "good" statistical response, that is, one that showed a correct use of the law of large numbers.

Such coding obviously runs into borderline cases. A coding guidebook was created which documented the principal types of borderline cases and the recommended treatment of them, for each problem. Reliability was tested by having four coders code a sample of 20 test booklets (300 law of large numbers problems). There was exact agreement among all four coders on 86% of these responses. Having achieved a high level of reliability, the primary coder (who had been one of the four coders), coded all of the responses, blind to conditions. His coding comprised the data we present here and in Experiment 2.

The coding of the three Structure 6 (false alarm) problems is described in a separate section below.

Training Procedures

All training procedures began with an introductory paragraph about decisions with limited information (quoted in full above as the first paragraph in the testing instructions for the control subjects).[1] Next followed a paragraph introducing the law of large numbers. This always began as follows:

> 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.*

Rule Training Condition

Subjects read a four-page description of the concept of sampling and the law of large numbers. This description introduced the important concepts associated with the law of large numbers and illustrated them by using the classic problem of estimating the true proportion of blue and red gumballs in an urn from a sample of the urn. Thus, the gumballs in the urn constituted the *population,* the proportion of blue and red gumballs in the urn formed the *population distribution* (in the example, the population distribution of gumballs was set at 70% blue and 30% red), and a selection of gumballs from the urn constituted a *sample.*

The concept of sampling was then presented by explaining that since it is often impractical or impossible to examine the entire population to determine the population distribution ("Imagine counting a million gumballs!"), it is necessary to rely instead on samples to *estimate* the population distribution. *Sample distributions,* subjects were told, vary in their closeness to the population distribution, and that the only factor determining the closeness of a *random* sample to the population is *sample size.* Finally, the law of large numbers was presented in the following way:

As the size of a random sample increases, the sample distribution is more likely to get closer and closer to the population distribution. In other words, the larger the sample, the better it is as an estimate of the population.

When subjects had finished reading this description, the experimenter performed a live demonstration of the law of large numbers, using a large glass urn filled with blue and red gumballs. In order to maximize subjects' understanding of the concepts they had just read, the demonstration was designed to adhere closely to the description. Each of the concepts introduced in the description was illustrated in the demonstration. For example, the population distribution of the urn was 70% blue and 30% red, just as it had been in the description.

After reintroducing all of the concepts, the experimenter drew four samples of size 1, then four of size 4, and finally, four of size 25. (The gumballs were returned to the urn after each sample.) The experimenter summarized each sample on a blackboard, keeping track of the deviation between each sample and the population. Subjects were told that the average deviation of a sample from the population would decrease as the sample size increased, in accordance with the law of large numbers. Thus, for example, samples of size 25 would, on the average, deviate less from the population than would samples of size 4 or 1. (By good luck, these expected results were obtained in all the training sessions.)

Examples Training Condition

Subjects in the examples training condition read a packet of three example problems with an answer following each problem that provided an analysis of it in terms of the law of large numbers. The three example problems were drawn from Structure 1 (generalizing from a small sample), Structure 3 (regression), and Structure 5 (large sample vs theory without supporting data), and were presented in that order. The three examples were all drawn from the domain of objective problems. After the paragraph that introduced the law of large numbers, there followed a single sentence describing one example of the principle (a public opinion poll based on a large sample is more likely to be accurate than one based on a small sample). The example problems were then introduced in the following way:

> The basic principles involved in the law of large numbers apply whenever you make a generalization or an inference from observing a sample of object, actions, or behaviors. To give you an idea of how broad the law of large numbers is, we have, in this packet, presented three situations in which the law of large numbers applies. Each situation is analyzed in terms of the law of large numbers.

For each example in turn, subjects read the problem and were asked to consider it for a few moments before turning the page to read the law of large numbers answer. The answers to the example problems were constructed so that subjects could learn how the law of large numbers might be applied to a variety of real-life situations. The format of the answers was constant across training domain and structure and included the following characteristics:

1. A statement about the goal of the problem;
2. Identification of the sample or samples and their distributions in the problem;
3. Explanation of how the law of large numbers could be applied to the problem. This identified the population distribution(s) and explained the relationship between the sample(s) and the population(s).
4. The conclusion that could be drawn from the application of the law of large numbers. The three example problems are presented in Appendix B.

Full Training Condition

Subjects received rule training, followed by examples training, except that the first sentence of the passage introducing the examples was replaced by the following sentence: "One reason that the law of large numbers is important to learn is that it applies *not only* to urns and gumballs."

Demand Condition

Subjects received only the one-sentence definition of the law of large numbers that introduced the examples training, along with the brief example. We included this condition in order to assess whether training effects might be due to experimenter demand or to simply making statistical rules salient to subjects. If performance of the demand group turned out not to be higher than that of the control group, these alternative explanations would be ruled out.

In addition, there was a *control* condition, which received no training before answering the test problems.

In summary, there were five conditions in Experiment 1, as shown in Fig. 5.1. They were defined by crossing the presence or absence of rule training with presence or absence of examples training. Note that the bottom-left cell of Fig. 5.1, where neither type of training was given, contains both the control and demand conditions.

Subjects and Procedure

The 347 subjects were adults (229) and high school students (118) from various New Jersey suburban communities. They were paid to participate in

		RULE TRAINING: NO	RULE TRAINING: YES
EXAMPLES TRAINING	YES	EXAMPLES	FULL TRAINING
EXAMPLES TRAINING	NO	DEMAND / CONTROL	RULE

FIG. 5.1 Design of Experiment 1.

the experiment. The adult subjects varied widely in age and education, but almost all were females who were not employed fulltime outside the home. Most of them had participated previously in psychology experiments at Bell Laboratories. Because adults and high school students showed the same pattern of results, their responses were combined in the analyses we present.

Subjects were scheduled in groups of 4–6, with the same training condition presented to the entire group. Training condition was randomly determined. Subjects were told the general nature of the experiment, given the appropriate training, and then given the 18-problem test booklet. They were given 80 min to complete the problems.

Results

Overview of Data Analysis

Recall that subjects' responses were coded using a 3-point system: A code of "1" was given for responses that contained no mention of statistical concepts such as variability or sample size, whereas a "2" or "3" was given for responses that incorporated statistical notions. Within the class of statistical responses, a "2" was given for "poor" statistical responses, and a "3" was given for "good" statistical responses.

We analyzed the data in terms of two dichotomies. The first one asks whether the response was deterministic (code = 1) or statistical, regardless of quality (code = 2 or 3). We refer to analyses based on this dichotomy as analyses of *frequency* of statistical responses. The second dichotomy asks, for statistical responses only, whether the response was poor (code = 2) or good (code = 3). We refer to analyses based on this dichotomy as analyses of *quality*. The quality dichotomy is conditional: it is defined only for statistical responses and is undefined (missing) for deterministic responses.

These two analyses allowed us to separate the questions of whether training increased the incidence of any kind of statistical reasoning from whether it increased the *proper* use of statistical principles. If we found that training led to an increase in frequency but a decrease in quality, this would lead to the pessimistic conclusion that training merely serves to make