

The Use of Statistical Heuristics in Everyday Inductive Reasoning

Richard E. Nisbett
University of Michigan

David H. Krantz
Bell Laboratories, Murray Hill, New Jersey

Christopher Jepson and Ziva Kunda
University of Michigan

In reasoning about everyday problems, people use statistical heuristics, that is, judgmental tools that are rough intuitive equivalents of statistical principles. Statistical heuristics have improved historically and they improve ontogenetically. Use of statistical heuristics is more likely when (a) the sample space and the sampling process are clear, (b) the role of chance in producing events is clear, or (c) the culture specifies statistical reasoning as normative for the events. Perhaps because statistical procedures are part of people's intuitive equipment to begin with, training in statistics has a marked impact on reasoning. Training increases both the likelihood that people will take a statistical approach to a given problem and the quality of the statistical solutions. These empirical findings have important normative implications.

It can be argued that inductive reasoning is our most important and ubiquitous problem-solving activity. Concept formation, generalization from instances, and prediction are all examples of inductive reasoning, that is, of passing from particular propositions to more general ones or of passing from particular propositions to other particular propositions via more general ones.

Inductive reasoning, to be correct, must satisfy certain statistical principles. Concepts should be discerned and applied with more confidence when they apply to a narrow range of clearly defined objects than when they apply to a broad range of diverse and loosely defined objects that can be confused with objects to which the concept does not apply. Generalizations should be more confident when they

are based on a larger number of instances, when the instances are an unbiased sample, and when the instances in question concern events of low variability rather than high variability. Predictions should be more confident when there is high correlation between the dimensions for which information is available and the dimensions about which the prediction is made, and, failing such a correlation, predictions should rely on the base rate or prior distribution for the events to be predicted.

Because inductive reasoning tasks are so basic, it is disturbing to learn that the heuristics people use in such tasks do not respect the required statistical principles. The seminal work of Kahneman and Tversky has shown that this is so and, also, that people consequently overlook statistical variables such as sample size, correlation, and base rate when they solve inductive reasoning problems. (See surveys by Einhorn & Hogarth, 1981; Hogarth, 1980; Kahneman, Slovic, & Tversky, 1982; Nisbett & Ross, 1980.)

The above research on *nonstatistical heuristics* has been criticized on several grounds. Some critics have maintained that evolution should be expected to produce highly efficacious and generally correct principles of reasoning and that the research may therefore be misleading in some way (Cohen, 1979; Dennett, 1978, 1981, Note 1; Lycan, 1981). Others

The research reported here was supported by Grants BNS79-14094 and BN575-23191 from the National Science Foundation. We are indebted to Lee Ross and Amos Tversky for their invaluable advice. We thank Geoffrey T. Fong for his assistance and criticism, and we thank Nancy Cantor, Robyn Dawes, Hillel Einhorn, Robin Hogarth, John Holland, Keith Holyoak, Nancy Pennington, Michael Shafto, Stephen Stich, and Paul Thagard for comments on an earlier version of the manuscript. Sara Freeland provided able editorial assistance.

Requests for reprints should be sent to Richard E. Nisbett, Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48106.

have maintained that the research does not demonstrate that people fail to apply correct inferential rules but rather that (a) it is the researchers themselves who are mistaken about the correct inferential rules (Cohen, 1981), (b) subjects have been misled by illusionary circumstances of little general significance beyond the laboratory (Cohen, 1981; Lopes, 1982; Dennett, Note 1), or (c) people's general inferential goals are such that at least some violations of statistical principles should be regarded as a form of *satisficing*, or cost-effective inferential shortcuts (Einhorn & Hogarth, 1981; Miller & Cantor, 1982; Nisbett & Ross, 1980).

We offer a different perspective on the incorporation of statistical principles into inductive reasoning, one that rejects the preceding criticisms but is, at the same time, fairly sanguine about people's statistical reasoning. Workers in the Kahneman and Tversky tradition have focused primarily on (a) establishing that people fail to respond to important statistical variables for a wide range of problems and (b) examining the inferential principles that people seem to rely on in solving such problems. There has been no comparable systematic effort to determine whether people do respond to statistical variables, either for problems that are easier than those examined to date or for problems of a different kind than those examined.

If it could be shown that people sometimes do reason using explicitly statistical principles, then the work to date on inductive reasoning, and the criticism of that work, would be cast in a different light. Rather than asking why the failures occur or whether the failures are real, it would seem more fruitful to ask questions such as the following. What factors encourage statistical reasoning and what factors discourage it? For what kinds of events and for what kinds of problems is statistical reasoning most likely to be used? Does purely formal training modify the untutored heuristics of everyday inductive reasoning? In addition, accusations that the work to date rests on a kind of experimental sleight of hand or that people are deliberately and advisedly setting aside statistical principles in favor of quicker and generally satisfactory procedures would seem less plausible. Instead, it would seem more likely that there are just difficul-

ties—surprisingly severe difficulties to be sure, but difficulties merely—in people's use of statistical principles for inductive reasoning.

In this article we first summarize the recent work establishing failures to reason statistically. We then review anecdotal and experimental evidence indicating that people do sometimes reason statistically. Next we present original experimental work indicating some of the factors that influence statistical reasoning. Then we summarize research suggesting that people's ability to reason statistically about everyday life problems is affected by training in formal statistics. Finally, we speculate on the normative implications of people's ability and trainability for statistical reasoning.

Statistical Problems and Nonstatistical Heuristics

In a succession of studies over the past decade, Kahneman and Tversky have shown that much inductive reasoning is nonstatistical. People often solve inductive problems by use of a variety of intuitive heuristics—rapid and more or less automatic judgmental rules of thumb. These include the representativeness heuristic (Kahneman & Tversky, 1972, 1973), the availability heuristic (Tversky & Kahneman, 1973), the anchoring heuristic (Tversky & Kahneman, 1974), and the simulation heuristic (Kahneman & Tversky, 1982). In problems where these heuristics diverge from the correct statistical approach, people commit serious errors of inference.

The representativeness heuristic is the best studied and probably the most important of the heuristics. People often rely on this heuristic when making likelihood judgments, for example, the likelihood that Object A belongs to Class B or the likelihood that Event A originates from Process B. Use of the heuristic entails basing such judgments on "the degree to which A is representative of B, that is, by the degree to which A resembles B" (Tversky & Kahneman, 1974, p. 1124). In one problem, for example, Kahneman and Tversky (1972) asked subjects whether days with 60% or more male births would be more common at a hospital with 15 births per day, or at a hospital with 45 births per day, or equally common at the two hospitals. Most subjects chose the latter alternative, and the remainder divided about

evenly between 15 and 45. The law of large numbers requires that, with a random variable such as sex of infant, deviant sample percentages should be less common as sample size increases. The representativeness heuristic, however, leads subjects to compare the similarities of the two sample proportions to the presumed population proportion (50%); because the two sample proportions equally resemble the population proportion, they are deemed equally likely. The data indicate that, for this problem at least, most subjects used the representativeness heuristic and very few subjects used the law of large numbers.

In another demonstration, Kahneman and Tversky (1973) studied the prediction of an outcome for a target person based on various characteristics of that person or based on scores from various predictor tests. Subjects used the representativeness heuristic: In general, they predicted whichever outcome was most similar to the target person's characteristics or scores. For instance, in predicting the grade point average (GPA) for a target person who is in the 90th percentile on a predictor test, about the same results are obtained—that is, prediction of a GPA well above average—whether the predictor is the score on a test of sense of humor (which subjects do not regard as very diagnostic of GPA), the score on a test of mental concentration, or the GPA itself (!). Such predictions diverge from those that would be obtained from statistical considerations in which the average accuracy of prediction would be taken into account. Subjects do not seem to realize that if accuracy is very limited, then it is far more probable that the target person's outcome will be equal to the modal outcome (or near the mean of the unimodal symmetric distribution) than that it will take some relatively unusual value that happens to match the characteristics on the predictor. This is the statistical principle of regression to the mean, or base rate.

Other investigations have confirmed and expanded the list of statistical failings documented by Kahneman and Tversky. The failings seem particularly clear and particularly important in people's reasoning about social behavior. Nisbett and Borgida (1975), for example, showed that consensus information, that is, base rate information about the behavior of a sample of people in a given situ-

ation, often has little effect on subjects' attributions about the causes of a particular target individual's behavior. When told that most people behaved in the same way as the target, subjects shift little or not at all in the direction of assuming that it was situational forces, rather than the target's personal dispositions or traits, that explain the target's behavior. In a typical experiment, Nisbett and Borgida (1975) told subjects about a study in which participants heard someone (whom the participants believed to be in a nearby room) having what seemed to be an epileptic seizure. Subjects' predictions about whether a particular participant would quickly help the "victim" were unaffected by the knowledge that most participants never helped or helped only after a long delay. Similarly, subjects' causal attributions about the behavior of a participant who never helped the "victim" were unaffected by consensus information. Subjects were just as likely to say that the participant's personality was responsible for his behavior when they knew that most other participants were similarly unhelpful as when they assumed that most other participants helped with alacrity.

Nisbett and Ross (1980) maintained that people fail to apply necessary statistical principles to a very wide range of social judgments. They claimed that people often make overconfident judgments about others based on small and unreliable amounts of information; they are often insensitive to the possibility that their samples of information about people may be highly biased; they are often poor at judging covariation between events of different classes (e.g., "Are redheads hot-tempered?"); and both their causal explanations for social events and their predictions of social outcomes are often little influenced by regression or base rate considerations.

Statistical Heuristics

Selective Application of Statistical Reasoning

The foregoing work indicates that nonstatistical heuristics play an important role in inductive reasoning. But it does not establish that other heuristics, based on statistical concepts, are absent from people's judgmental repertoire. And indeed, if one begins to look for cases of good statistical intuitions in ev-

eryday problems, it is not hard to find some plausible candidates.

Even when judgments are based on the representativeness heuristic, there may be an underlying stratum of probabilistic thinking. In many of the problems studied by Kahneman and Tversky, people probably conceive of the underlying process as random, but they lack a means of making use of their intuitions about randomness and they fall back on representativeness. In the maternity ward problem, for example, people surely believe that the number of boys born on any particular day is a matter of chance, even though they rely on representativeness to generate their subjective sampling distributions. But consider the following thought experiment: If someone says, "I can't understand it; I have nine grandchildren and all of them are boys," the statement sounds quite sensible. The hearer is likely to agree that a causal explanation seems to be called for. On the other hand, imagine that the speaker says, "I can't understand it; I have three grandchildren and all of them are boys." Such a statement sounds peculiar, to say the least, because it seems transparent that such a result could be due just to chance—that is, there is nothing to understand. Such an intuition is properly regarded as statistical in our view.

The contrast between the statistical intuition in our anecdote and subjects' use of the representativeness heuristic in the maternity ward problem illustrates the selectivity with which people apply statistical concepts. The failure to do so in the maternity ward problem may be due to the use of "60%" in the problem, which evokes comparison between 60% and 50% and thence the dependence on the similarity judgment in choosing an answer. It may also be due to lack of concrete experience in thinking about samples in the range 15–45. As Piaget and Inhelder (1951/1975) put it, people seem to have an intuitive grasp of the "law of *small* large numbers," even though they may not generalize the intuition to large numbers.

People also seem to have an ability to use base rates for selected kinds of problems. Consider the concepts of *easy* and *difficult* examinations. People do not infer that a student is brilliant who received an A+ on an exam in which no one scored below A— nor that

the student is in trouble who flunked a test that was also failed by 75% of the class. Rather, they convert the base rate information (performance of the class as a whole) into a location parameter for the examination (easy, . . . , difficult) and make their inference about the particular student in terms of the student's relative position compared to the difficulty of the exam. Indeed, laboratory evidence has been available for some time that base rates are readily utilized for causal attributions for many kinds of abilities and achievements (Weiner et al. 1972).

As Nisbett and Ross (1980) suggested, one suspects that many lay concepts and maxims reflect an appreciation of statistical principles. It seems possible, for example, that people sometimes overcome sample bias by applying proverbs such as "Don't judge a book by its cover" or "All that glitters is not gold." Perhaps people sometimes even manage to be regressive in everyday predictions by using concepts such as "beginner's luck" or "nowhere to go but up/down."

There is one inductive reasoning task in particular for which there is good reason to suspect that statistical intuitions are very frequently applied. This is *generalization from instances*—perhaps the simplest and most pervasive of everyday inductive tasks. People surely recognize, in many contexts at least, that when moving from particular observations to general propositions, more evidence is better than less. The preference for more evidence seems well understood as being due to an intuitive appreciation of the law of large numbers. For example, we think that most people would prefer to hold a 20-minute interview rather than a 5-minute interview with a prospective employee and that if questioned they would justify this preference by saying that 5 minutes is too short a period to get an accurate idea of what the job candidate is like. That is, they believe that there is a greater chance of substantial error with the smaller sample. Similarly, most people would believe the result of a survey of 100 people more than they would believe that of a survey of 10 people; again, their reason would be based on the law of large numbers.

As we shall see, there is reason to believe that people's statistical understanding of the generalization task is deeper still. People un-

derstand, at least in some contexts, that the law of large numbers must be taken into account to the degree that the events in question are uncertain and variable in a statistical sense. Thus they realize that some classes of events are very heterogeneous; that is, the events differ from one another, or from one occasion to another, in ways that are unpredictable, and it is these classes of events for which a large sample is particularly essential.

Randomizing Devices and the Ontogeny of Statistical Reasoning

Where do people's selective statistical intuitions come from? An extremely important series of studies by Piaget and Inhelder (1951/1975) suggests that the intuitions may arise in part from people's understanding of the behavior of random generating devices. Statistical reasoning is of course very commonly applied in our culture to the behavior of such mechanisms. Piaget and Inhelder showed that statistical intuitions about random devices develop at an early age. They conducted experiments in which children were shown various random generating devices and then were asked questions about them. The devices included different-colored marbles on a tilt board, coin tosses, card draws, a spinner, and balls dropped through a funnel into a box with a varying number of slots. Children were shown the operation of these devices and then were asked to predict outcomes of the next operation or set of operations and to explain why particular outcomes had occurred or could or could not occur. The work showed that even children less than 10 years old used the concept of chance and understood the importance of sequences of repeated trials.

In one study, for example, Piaget and Inhelder (1951/1975) spun a pointer that could stop on one of eight different-colored locations. The young children they studied (in general, those less than 7 years old) did not initially recognize their complete inability to predict the pointer's stopping place.

He knows quite well that he is not likely to be able to predict the color on which the bar will stop, but he does believe in the legitimacy of such a prediction and tries to guess the result. . . . The child oscillates quickly between two solutions . . . Either the bar will have the tendency to come back to a color on which it has already stopped,

or it will, on the contrary, stop on the colors not yet touched (p. 61).

At this stage the children did not recognize the equivalent chances of the various stopping places, and when the pointer was made to stop at one color repeatedly (by using a magnet) they found nothing unusual in this. A satisfactory causal explanation usually was forthcoming: for example, "the pointer got tired."

By around the age of 7, the Piaget and Inhelder subjects began to understand the chance nature of the pointer's behavior. After a few demonstrations, they quickly came to doubt the predictability of single trials and came to see the distribution of possibilities and their equivalence. Between the ages of 7 and 10, their subjects came to understand the importance of repeated trials and long run outcomes.

E: If I spun it ten or twenty times, could there be one color at which it never stopped?

S (age 7): Yes, that could happen. That would happen more often if we did it only ten times rather than twenty (p. 75).

E: Will it hit all the colors or not?

S (age 10 years, 7 months): It depends on how long we spin it.

E: Why?

S: Because if we spin it often, it will have more chances of going everywhere (p. 89).

How does the child come to have an understanding of the concept of chance during this period? Piaget and Inhelder argue that *the child's understanding of uncertainty grows out of the child's understanding of physical causality*. To the very young child with little understanding of the causal mechanisms that produce outcomes in a physical system, every outcome is a "miracle"—that is, unanticipated—and, paradoxically, once the outcome has occurred, the child believes that it can be explained. As the child comes to understand, in terms of concrete operations, the causal mechanisms that produce outcomes, the child begins to recognize which sorts of outcomes are predictable (and explainable) and which are not. The outcomes that are not predictable are gradually understood to obey certain *non-causal rules*. In particular, the child comes to recognize some cases of the law of large numbers, for example, that the likelihood of any given outcome occurring is greater with a large number of trials than with a smaller number.

By the age of 11 or so, many children have—in addition to a clear conception both of fully deterministic systems and of random generating devices—a good understanding of non-uniform probability distributions. These are partially random systems in which causal factors are at work making some of the possible outcomes more likely than others. The child comes to learn that even though individual events are uncertain in such a system, aggregate events may be highly predictable. In such a probabilistic system, the child grasps the relevance to prediction of the base rate, that is, the distribution and relative frequency of the various outcomes.

This latter point is well illustrated by children's understanding of a device that allows balls to be dropped through a hole into one of a number of slots or bins beneath. Here the chances of a ball dropping into one slot versus another can be made quite unequal by the physical set-up. It is easy to build the device, for example, so that most balls drop into middle bins and fewer drop in the side bins, generating a crude bell curve. Children under 7 generally fail to use this distribution as a basis of prediction. Although they slowly come to recognize that central positions will collect more balls than peripheral ones, they cannot generalize this fact from a box with a particular number of slots to another box with a different number; they do not expect symmetry between slots that are equidistant from the center; and they do not recognize the role of the law of large numbers in making the central slots particularly favored over a long series of trials. All of these intuitions, in contrast, come easily to many 12-year-olds.

We may speculate that the older child's statistical conceptualization of the behavior of randomizing devices serves as the basis for a similar conceptualization of other kinds of events that may be seen as variable and uncertain. We discuss later just what characterizes events where an analogy to randomizing devices can be seen versus those where it cannot be seen.

The Intellectual History of Statistical Reasoning

The cultural history of statistical reasoning appears to parallel in some interesting respects the developmental course described by Piaget and Inhelder (1951/1975). This history has

been traced by Hacking in his book *The Emergence of Probability* (1975). Hacking points out that although random generating devices have been used at least since Biblical times, the modern concept of probability was invented rather suddenly in the 17th century. This was true despite the popularity of games of chance in antiquity and the existence of sophisticated mathematics. (Hacking notes that someone with only a modest knowledge of modern probability could have won all Gaul in a week!)

Paradoxically, the major factor underlying the sudden emergence of the modern concept of probability was the change to a deterministic understanding of the physical world. In the Renaissance, the task of science was understood not primarily as a search for the causal factors influencing events but as a search for signs as to the meaning of events. These signs were clues and portents strewn about by the benign Author of the Universe. This sort of understanding of events encouraged a heavy reliance on the representativeness heuristic. The Renaissance physician, for example, adhered to the *doctrine of signatures*. This was the "belief that every natural substance which possesses any medicinal virtue indicates by an obvious and well-marked external character the disease for which it is a remedy, or the object for which it should be employed" (John Paris, cited in Mill, 1843/1974, p. 766). The representativeness heuristic thus could be derived as a rule of inference from the principle that the Author of the Universe wanted to be helpful in our attempts to understand the world.

A quite different way of understanding events became predominant in the 17th century. This was a new "mechanistic attitude toward causation" (Hacking, 1975, p. 3). Just as the development of concrete operations helps the child to recognize the irreducible ignorance and uncertainty that is left as a residue after causal analysis of a randomizing device, so the new attitude toward causation helped 17th century scientists appreciate the nature of uncertainty in probabilistic systems. "Far from the 'mechanical' determinism precluding an investigation of chance, it was its accompaniment . . . this specific mode of determinism is essential to the formation of concepts of chance and probability" (Hacking, 1975, p. 3).

Summary

In short, there is good reason to believe that people possess *statistical heuristics*—intuitive, rule-of-thumb inferential procedures that resemble formal statistical procedures. People apply these heuristics to the behavior of random generating devices at a fairly early age. The formal understanding of statistical principles—that is, of the rules governing the behavior of randomizing devices—increases at least until adolescence. The use of such heuristics, both individually and culturally, seems related to the growth of causal understanding of the physical world and to attempts to extend this causal understanding, by analogy, to wider domains. Although we know little at present of the growth in the child's or adolescent's ability to apply statistical heuristics to events other than those produced by randomizing devices, it seems clear that such growth does take place. Adults who are untutored in formal statistics seem to reason statistically about a number of events other than those produced by randomizing machines—such as performance on tests, sports, weather, and accident and death risks. In addition, it is hard to imagine that people could conduct the most basic of inferential tasks, namely, generalization from instances, without the application of at least a rudimentary version of a law-of-large-numbers heuristic.

Factors That Affect Statistical Reasoning

Despite ontogenetic and historical growth in the ability to reason statistically, contemporary adults do not reason statistically about a wide range of problems and event domains that require such reasoning, and they often do not do so even if they have substantial training in formal statistics (Tversky & Kahneman, 1971). Why is this? What factors make it difficult to apply statistical heuristics when these are required, and what factors can make it easier? Three factors that seem important are implicit in the preceding discussion.

Clarity of the Sample Space and the Sampling Process

Randomizing devices are usually designed so that the sample space for a single trial is obvious and so that the repeatability of trials is salient. The die has six faces and can be tossed again and again; the pointer can stop

on any of eight sectors and can be spun over and over. Clarity of sample space makes it easier to see what knowledge is relevant. For randomizing devices, the most relevant knowledge is often just the observation of symmetry of the different die faces, spinner sectors, and so forth. The salience of repeatability makes it easier to conceptualize one's observations as a sample.

In the social domain, sample spaces are often obscure, and repeatability is hard to imagine. For example, the sample space consisting of different degrees of helpfulness that might be displayed by a particular person in a particular situation is quite obscure, and the notion of repetition is strained. What is it that could be repeated? Placing the same person in *different* situations? Or *other* people in the same situation? The probability that Person P will exhibit Behavior B in Situation S is abstract and not part of the inductive repertoire of most people most of the time. Even though people recognize the possibility of errors in their judgments of social situations, they do not try to construct probability models; rather, they rely on the representativeness heuristic.

Recognition of the Operation of Chance Factors

A second major factor encouraging the use of statistical heuristics is the recognition of the role of chance in producing events in a given domain or in a particular situation. We have already seen how Piaget and Inhelder (1951/1975) describe the recognition of chance in the operation of randomizing devices. The child comes to recognize the limitations of causal analysis for a spinner and the consequent residual uncertainty about the production of events. Something like the same transparent indeterminism exists for other sorts of events as well, even those involving human beings. For example, statistical understanding of some types of sports is undoubtedly facilitated by the manifestly random component in the movement of the objects employed: "A football can take funny bounces." The random component probably does not have to be physical in order for people to recognize it. It is possible to recognize the unpredictability of academic test performance by repeated observations of one's own outcomes. Even with

one's own efforts and the group against which one is competing held constant, outcomes can vary. One may even recognize that one's performance on particular occasions was particularly good or poor because of accidents: "I just happened to reread that section because Jill never called me back"; "It was very noisy in the study area that night so I didn't get a chance to review my notes."

In contrast, cues as to randomness in the production of events are much subtler for other kinds of events, especially for many social ones. When we interview someone, what signs would let us know that a particular topic got explored just by chance or that the person seems dour and lackluster because of an uncharacteristic attempt to appear dignified rather than because of a phlegmatic disposition? In addition, as Einhorn and Hogarth (1978) have pointed out, the gatekeeping function of the interview may serve to prevent us from recognizing the error variance in our judgments: The great talent of some people not hired or admitted may never be observed. Daniel Kahneman (Note 2) has suggested to us that the "interview illusion" exists in part because we expect that brief encounters with a living, breathing person ought to provide a "hologram" of that person rather than merely a sample of the person's attributes and behaviors. In most situations, cues as to the fact that an interview ought to be regarded as a sample from a population, rather than a portrait in miniature, are missing. The same may be true for visits to a city, country, or university. One of us long believed that reports of raininess in England were greatly exaggerated because he once stayed in London for 10 days and it only drizzled twice!

Cultural Prescriptions

A third factor that may contribute to the use of statistical heuristics is a cultural or sub-cultural prescription to reason statistically about events of a given kind. Although Piaget and Inhelder focused on developmental changes in the ability to reason statistically about randomizing devices, from a historical perspective it is the young child's ability to reason statistically at all about such devices that is remarkable. It seems implausible that a medieval European child would have reasoned in such a sophisticated way as the Piaget

and Inhelder subjects. Statistical reasoning is the culturally prescribed way to think about randomizing devices in our culture, and this general approach undoubtedly trickles down to children. Similarly, statistical reasoning has become (or is becoming) the norm for experts in many fields—from insurance to medical diagnosis—and is rapidly becoming normative for the lay novice as well in such domains as sports and the weather. Models of statistical reasoning abound for sports in particular, as the two examples below indicate.

Baseball's law of averages is nothing more than an acknowledgement that players level off from season to season to their true ability—reflected by their lifetime averages. A .250-hitter may hit .200 or .300 over a given period of time but baseball history shows he will eventually level off at his own ability ("Law of Averages," 1981).

The musky tends to be a deep water fish. Most fishing success is in shallow water, but . . . this misleading statistic [is probably accounted for in part by the fact that] sheer statistical chance dictates that fish will come from the waters receiving the most man hours of fishing pressure. Shallow water fishing for muskies is very popular, and very few fishermen work them deep (Hamer, 1981).

The statistical spirit embodied in these quotations reaches many fans. Thus, it is commonplace to hear lay people endorse the proposition that "On a given Sunday any team in the NFL can beat any other team." (Compare with "On a given Sunday, any parishioner's altruism can exceed that of any other parishioner"!)

In our view, these three factors—clarity of the sample space and the sampling process, recognition of the role of chance in producing events, and cultural prescriptions to think statistically—operate individually and, perhaps more often, together to increase people's tendencies to apply statistical heuristics to problems that require a statistical approach. If these factors are genuinely important determinants of people's ability to reason statistically, then it should be possible to find support for the following predictions.

In cases where the sample space is clear and the possibility of repetition is salient, people will respond appropriately to statistical variables. In particular, in the task of generalizing from instances, where the sample space is a clear dichotomy and the sampling process is just the observation of more members of a clearly defined population, (a) people will generalize more cautiously when the sample size

is small and when they have no strong prior belief that the sampled population is homogeneous, and (b) people can be influenced to generalize more or less readily by manipulations that emphasize the homogeneity or heterogeneity of the sampled population.

The following predictions should hold both for generalization and for other, more complex, inferential tasks: (a) Manipulations designed to encourage recognition of the chance factors influencing events should serve to increase statistical reasoning. (b) People who are highly knowledgeable about events of a given kind should be more inclined than less knowledgeable people to apply statistical reasoning to the events—because both the distributions of the events and the chance factors influencing the events should be clearer to such people. (c) People should be disinclined to reason statistically about certain kinds of events that they recognize to be highly variable and uncertain—notably social events—because the sample spaces for the events and the chance factors influencing the events are opaque. (d) Training in statistics should promote statistical reasoning even about mundane events of everyday life because such training should help people to construct distributional models for events and help them to recognize “error,” or the chance factors influencing events.

Generalizing From Instances

Generalization from observed cases is the classic concern of philosophers and other thinkers who are interested in induction. A number of instances of Class A are observed, and each of them turns out to have Property B. Possible inferences include the universal generalization *all A's have B*, or the near universal *most A's have B*, or at least the relinquishing of the contrary generalization, namely, *most A's do not have B*.

The untrammelled employment of the representativeness heuristic would lead people to make the above inferences from quite small numbers of instances, and, indeed, this is often found, both anecdotally and in laboratory studies (Nisbett & Ross, 1980, pp. 77–82). On the other hand, philosophers since Hume have puzzled about how these generalizations can be logically justified, even when very large numbers of instances are observed. The puzzle

has been compounded by the fact that sometimes it seems correct to generalize confidently from a few instances. Hume (1748/1955) wrote, “[Often, when] I have found that . . . an object has always been attended with . . . an effect . . . I foresee that other objects which are in appearance similar will be attended with similar effects” (p. 48). The problem is that only sometimes do we draw such a conclusion with confidence. “Nothing so like as eggs, yet no one, on account of this appearing similarity, expects the same taste and relish in all of them” (p. 50). Mill (1843/1974), a century later, phrased the problem like this: “Why is a single instance, in some cases, sufficient for a complete induction, while in others myriads of concurring instances, without a single exception known or presumed, go such a very little way towards establishing a universal proposition?” (p. 314).

The statistical advances since Mill’s time make it clear that a large part of the answer to his question has to do with beliefs about the variability or *homogeneity* of certain kinds or classes of events (cf. Thagard & Nisbett, 1982). Generalization from a large sample is justified in terms of one’s beliefs that the sampling itself is homogeneous (i.e., that the distribution of possible sample statistics is the same as would be predicted by random sampling). And generalization from a small sample or resistance to generalization, even from a large sample, are justified in terms of prior beliefs about the homogeneity or heterogeneity of objects or events of a certain kind with respect to a property of a certain kind. If, for example, the object is one of the chemical elements and the property is electrical conductivity, then one expects homogeneity: All samples of the element conduct electricity or none do. But if the object is an animal and the property is blueness, one’s prior belief does not favor homogeneity so strongly; color may or may not vary within a particular species.

In other words, there are cases where use of the representativeness heuristic is justified in terms of beliefs about homogeneity, which in turn may be soundly based on individually or culturally acquired experience with kinds of objects and kinds of properties. For other cases, simple representativeness cannot be justified, and there are indeed cases, as Mill claimed, in which a strong prior belief in *het-*

erogeneity properly prevents acceptance of a generalization even after quite large numbers of instances have been observed.

We attempted to demonstrate, in a laboratory study of judgment, that people do in fact temper the use of representativeness to a greater or lesser degree depending on beliefs about the variability of a kind of object with respect to a kind of property.

Study 1: Beliefs About Homogeneity and Reliance on the Law of Large Numbers

In this study, we simply guessed at the prevailing beliefs about homogeneity. We tried to obtain different degrees of heterogeneity by using conductivity of metals, colors of animals, and so on. Subjects were told of one instance or of several instances of a sampled object having a particular property and were asked to guess what percentage of the population of all such objects would have the property. The sample sizes used were 1, 3, or 20; in the latter cases, all 3 or all 20 of the objects had the property in question. We anticipated that subjects would generalize more readily from a given number of instances when the kind of object was perceived as homogeneous with respect to the kind of property than when the kind of object was perceived as heterogeneous with respect to the kind of property.

Method

Subjects were 46 University of Michigan students of both sexes who were enrolled in introductory psychology. (As sex did not affect any of the dependent variables in this or any of the other studies, it will not be discussed further.) Eighty-five percent of the subjects had taken no statistics courses in college. The questionnaire was presented as one of several in a study on judgment. It read as follows for the $N = 1$ condition:

Imagine that you are an explorer who has landed on a little known island in the Southeastern Pacific. You encounter several new animals, people, and objects. You observe the properties of your "samples" and you need to make guesses about how common these properties would be in other animals, people or objects of the same type.

Suppose you encounter a new bird, the shreeble. It is blue in color. What percent of all shreebles on the island do you expect to be blue?

(This and the subsequent questions were followed by

"___ percent. Why did you guess this percent?")

Suppose the shreeble you encounter is found to nest in

a eucalyptus tree, a type of tree which is fairly common on the island. What percent of all shreebles on the island do you expect to nest in eucalyptus trees?

Suppose you encounter a native, who is a member of a tribe he calls the Barratos. He is obese. What percent of the male Barratos do you expect to be obese?

Suppose the Barratos man is brown in color. What percent of male Barratos do you expect to be brown (as opposed to red, yellow, black or white)?

Suppose you encounter what the physicist on your expedition describes as an extremely rare element called floridium. Upon being heated to a very high temperature, it burns with a green flame. What percent of all samples of floridium found on the island do you expect to burn with a green flame?

Suppose the sample of floridium, when drawn into a filament, is found to conduct electricity. What percent of all samples of floridium found on the island do you expect to conduct electricity?

The questionnaires for the $N = 3$ condition and the $N = 20$ condition were identical except that they specified larger samples of each object. For example, the first shreeble item for the $N = 3$ condition read as follows:

Suppose you encounter a new bird, the shreeble. You see three such birds. They are all blue in color. What percent of all shreebles on the island do you expect to be blue?

The reasons subjects gave for guessing as they did were coded as to their content. There were three basic sorts of answers: (a) references to the homogeneity of the kind of object with respect to the kind of property, (b) references to the heterogeneity of the kind of object with respect to the kind of property—due to the different properties of subkinds (e.g., male vs. female), to some causal mechanism producing different properties (e.g., genetic mistakes), or to purely statistical variability (e.g., "where birds nest is sometimes just a matter of chance"), and (c) other sorts of answers that were mostly based on representativeness or that were mere tautologies. Two independent coders achieved 89% exact agreement on coding category.

Results

Any one element is presumed by scientists to be homogeneous with respect to most properties. At the other extreme, most human groups are highly heterogeneous among themselves in many attributes, including body weight. If educated lay people share these beliefs and if they reason statistically, then (a) they should exercise more caution in generalizing from single cases when heterogeneity is expected than when homogeneity is expected and (b) large N should be important primarily in the case of populations whom subjects believe to be heterogeneous with respect to the property in question.

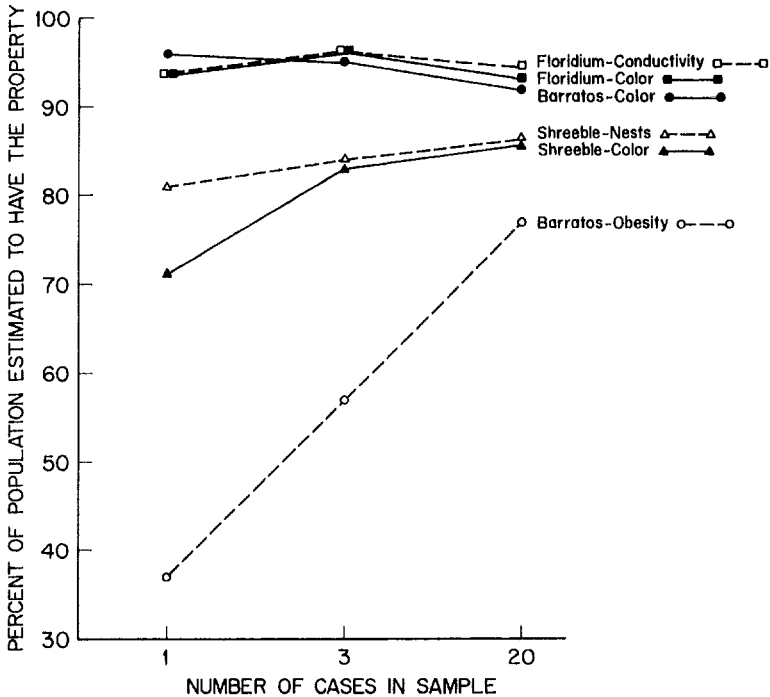


Figure 1. Percentage of each population estimated to have the sample property as a function of number of cases in the sample.

Figure 1 presents subjects' estimates of the percentage of each population having the property associated with the sample as a function of sample size presented. It may be seen that subjects are quite willing to generalize from even a single instance of green-burning or electricity-conducting floridium and also from a single, brown, Barratos tribesman. The modal estimate for $N = 1$ (as well as for $N = 3$ and $N = 20$) in all of these cases is 100%. In contrast, generalizations are less extreme for even 20 instances of blue shreebles or eucalyptus-nesting shreebles or 20 obese Barratos. The $t(31)$ contrasting $N = 1$ for floridium attributes and Barratos color with $N = 20$ for shreeble attributes and Barratos obesity is 3.00; $p < .01$.¹

Subjects' explanations for their estimates fully justify this pattern of inferences. It may be seen in Table 1 that subjects reported believing that elements are homogeneous with respect to color and conductivity and that tribes are homogeneous with respect to color. In contrast, subjects rarely expressed the belief that there is homogeneity for the other kinds

of populations and properties and instead expressed belief in heterogeneity of one sort or another for these objects and properties. Figure 1 shows that it is only for these latter cases that subjects reasoned statistically in the sense that they were more willing to assume that the population resembles the sample when N is larger. N affects the estimates of the obesity of Barratos and the color of shreebles ($p < .001$ and $p = .11$, respectively). In addition, a total of 10 subjects complained on one or more problems that the N was too small to give a good estimate. For nine of these subjects, the complaints were about one or more of the three problems where populations were presumed to be heterogeneous with respect to the property in question, whereas for only one subject was the complaint about a problem for which subjects in general believed populations to be homogeneous with respect to properties (exact $p = .02$).

¹ All p values are based on two-tailed tests unless otherwise indicated.

Finally, an internal analysis of the Table 1 data for each question showed that those subjects who believed the population to be homogeneous with respect to the property estimated that a higher percentage of the population was like the sample than did those subjects who believed the population was heterogeneous with respect to the property. The lowest t resulting from the six comparisons yielded $p < .05$.

Study 2: Manipulating the Salience of Distribution Parameters

Study 1 established that people can apply statistical reasoning to one of the most basic of inferential tasks. It also established that beliefs about variability of the class of events in question can mediate the statistical reasoning. One other study in the literature made similar points. Quattrone and Jones (1980) proposed a version of the present view that beliefs about variability influence inductive generalizations in their important study on perception of ingroups versus outgroups. They hypothesized that "an observer's tendency to generalize from the behavior of a specific group member to the group as a whole is proportional to the observer's perception of the group's homogeneity" (p. 141). Because people are more familiar with the members of groups to which they happen to belong, they will recognize "the group's general variability, the extent to which its members . . . differ from one another when

viewed over all dimensions" (p. 141). Because people are less familiar with outgroups, they are at liberty to assume that their members are relatively uniform. Thus people may generalize more readily from observations of the behavior of outgroup members than from observations of the behavior of ingroup members.

To test this hypothesis, Quattrone and Jones (1980) showed Princeton and Rutgers University undergraduates videotapes of male students who were allegedly serving as participants in psychology experiments. These students were asked to make choices such as to wait for a few minutes by themselves versus in the company of others or to listen to rock music versus classical music. Half of the subjects at each campus believed they were viewing Princeton men, and half believed they were viewing Rutgers men. After seeing the choice of one participant, subjects were asked to predict what the 100 participants in the study did. Quattrone and Jones found greater generalization from the participants' behavior to outgroup members than to ingroup members. Thus, Princeton subjects generalized more strongly to the behavior of the Rutgers population after observing the choice of the "Rutgers" participant than they did to the Princeton population after observing the choice of the "Princeton" participant.

If, as both we and Quattrone and Jones assume, generalizations about groups from the behavior of its members are mediated by assumptions about variability of group members,

Table 1
Number of Subjects Giving Each Type of Reason and Percentage of Population Estimated to Have the Property

Object and property	Reason					
	Homogeneity		Tautology		Heterogeneity	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Shreeble						
Color	6	95	17	83	22	75
Nests	8	96	19	84	19	78
Barratos						
Obesity	5	79	10	62	31	53
Color	31	98	7	94	8	80
Floridium						
Color	31	97	9	91	6	82
Conductivity	31	98	7	92	8	82

then it should be possible to manipulate those assumptions and therefore to influence the degree of generalization. People are inclined to think of (their own) university populations as being immensely variable—what with caftans here and exotic accents there, football players here and budding physicists there. In fact, however, university populations are not as heterogeneous as one might casually presume. Most students, even at multiversities, are, after all, bright young middle-class people of fairly homogeneous geographic and ethnic backgrounds. It seems possible that, if subjects were required to contemplate the central tendencies of their university populations before observing choice behavior like that presented to Quattrone and Jones's subjects, they might generalize more. This possibility was examined in Study 2.

Method

The procedure used by Quattrone and Jones (1980) was followed almost exactly, except that subjects were told that the videotapes were either of University of Michigan or of Ohio State University students, and half of the subjects were exposed to a central-tendency manipulation before viewing the videotapes. Subjects were 115 University of Michigan undergraduates of both sexes enrolled in introductory psychology. They participated in small groups, seated around a table facing a .53-m (21-inch) video monitor. Subjects were told that the investigators were "studying how people make judgments about people—working from actual information they have about people to guesses about other aspects of people. One of our major interests is in how students perceive students at (their own/another) university."

At this point the central-tendency manipulation was delivered to experimental subjects, who were told that "we will be asking you several questions about students at (the University of Michigan/Ohio State University)" and were given the appropriate central-tendency questionnaire. Control subjects began viewing videotapes immediately.

The central-tendency questionnaire consisted of three questions that we expected would influence subjects' conceptions of the variability of a student population. Subjects were asked to "please list what you would guess to be the 10 most common majors at (the University of Michigan/Ohio State University)" and next to list the five most common ethnic group backgrounds and the five most common religious backgrounds at that university. Answering these questions might be expected to prompt subjects to recognize that the student body is not all that heterogeneous: Most students are, after all, white Protestants concentrated in a limited number of relatively popular majors.

Subjects viewed the Quattrone and Jones videotapes.² They were introduced as having been made during psychology experiments conducted at the University of Michigan or at Ohio State University. In each of the three tapes

a male participant was shown being confronted with a decision, and he then chose one of two alternative behaviors offered. In the first scenario, a target person had to choose between waiting alone or waiting with other subjects while his experimenter fixed a machine. In the second scenario, the choice was between listening to classical music or listening to rock music during an experiment on auditory perceptual sensitivity. In the third scenario, the choice was between solving mathematical problems or solving verbal problems during an experiment on the effects of noise on intellectual performance. As the order in which scenarios were presented had no effect in the Quattrone and Jones study, it was held constant in our study.

The procedure was the same for each scenario. Subjects watched the target person being given instructions and being asked to make his decision. At this point the tape was turned off and subjects were asked to predict the target person's decision on a 21-point scale that had endpoints labeled with the two relevant options. The tape was then turned on again and subjects observed the participant make his decision. Half of the subjects saw the participants in the three scenarios make one set of decisions, and half saw the complementary set. Thus, subjects in Set A saw the target persons choose (a) to wait alone, (b) to listen to classical music, and (c) to solve mathematical problems. Subjects in Set B saw targets choose (a) to wait with others, (b) to listen to rock music, and (c) to solve verbal problems.

The dependent variable of interest consisted of the subjects' estimates of how many out of 100 participants in each of the three experiments chose each of the two options. (For the sole purpose of replication, subjects were also asked to indicate what they would have done and who they liked as people more—those who would prefer Option A or those who would prefer Option B.)

Results

Figure 2 presents subjects' generalizations about the University of Michigan and Ohio State University populations for control subjects and for subjects exposed to the central-tendency manipulation. *Generalization* is defined as the difference between population estimates for subjects presented with Set A choices versus those for subjects presented with Set B choices. The higher this index is, the more a group of subjects was influenced in their estimates by the behavior of the particular subject they witnessed. The index sums across all three types of choices, but the trends were the same for each of the three problems.

The difference between the control groups exposed to Ohio State University participants versus those exposed to University of Michigan participants provides a replicaton of the Quattrone and Jones finding. The magnitude of the

² We are indebted to George Quattrone for making these available.

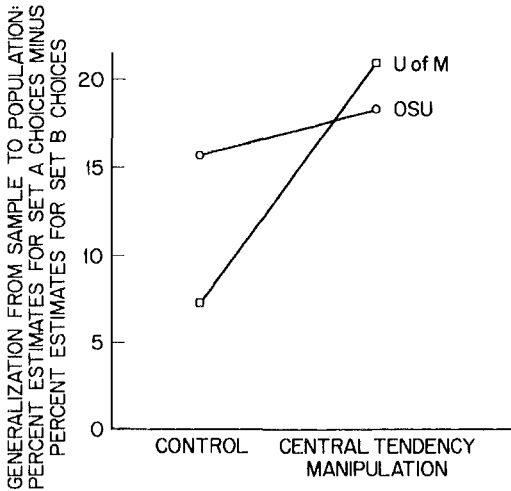


Figure 2. Generalization from sample to population as a function of campus population and central-tendency manipulation. (U of M = University of Michigan; OSU = Ohio State University.)

difference is very similar to that found by them, though for our smaller sample it is only marginally significant, $F(1, 50) = 2.76$, $.05 < p < .10$.

The effect of the central-tendency manipulation was to increase the degree of generalization from the sample, $F(1, 107) = 4.23$, $p < .05$. It may be seen that the effect was largely due to the behavior of the University of Michigan group. This is not surprising because the judgments about the Ohio State students may have already incorporated central tendencies in the form of an outgroup stereotype. This explanation should be viewed with caution, however, inasmuch as the interaction failed to reach statistical significance.

Both findings provide support for the contention that concurrent representations of population variability mediate inductive generalizations. Familiarity with one's own group results in less willingness to generalize for them than for another group, although forced contemplation of central tendencies results in more willingness to generalize, at least for the familiar ingroup.

One other study, by Silka (1981), shows the importance for inductive reasoning of people's focus on variability versus central tendency. She asked subjects to examine a series of numerical values that were said to represent the mental health of several individuals. Some

subjects were asked to remember the average of the values, and some were asked to remember the range. When subjects were asked, 1 week later, to assess the degree of change represented by a new value, subjects who had been asked to remember the average were more likely to infer that there had been a genuine change than those who had been asked to remember the range. The implication of Silka's finding, together with those of Study 2, is that inferences about continuity and change, and inductive reasoning generally, may be in part a function of arbitrary encoding and retrieval factors that accidentally emphasize either the homogeneity or the heterogeneity of events.

Study 3: Manipulating the Salience of Chance Factors

Study 2 establishes that manipulations of the salience of distributional parameters can influence subsequent generalizations. It should also be possible to influence generalizations by manipulating the salience of chance factors. One potentially interesting way of doing this would be to highlight for subjects the degree to which evidence about an object should properly be regarded as a sample from the population of the object's attributes. Such a reminder ought to prompt subjects to reason more statistically, deemphasizing evidence from smaller samples and placing greater weight on evidence from larger samples.

Borgida and Nisbett (1977) argued that people often ignore the judgments of others when choosing between two objects and substitute their own initial impressions of the objects as the sole basis of choice. People do this in part because they do not recognize the relevance of the law of large numbers when reasoning about events of the personal preference kind. When the objects are multifaceted and complex, however, the law of large numbers is applicable in two ways: (a) The reactions of other people to the object, especially if they are based on more extensive contact with the object than one has had oneself, generally should be a useful guide to choice (though, of course, it is possible to construct cases where other people's reactions would not be useful). (b) One's own experience with the object, especially if it is brief or superficial, may be a poor guide to choice because of the error

that plagues any small samples, even those that happen to be our own.

It seemed likely that if people were made explicitly aware of the role of chance in determining the impression one may get from a small sample, they might place less faith in a small personal sample and more faith in a large sample based on other people's reactions.

Method

Subjects were 157 University of Michigan students of both sexes who were enrolled in introductory psychology classes. Eighty-seven percent had taken no statistics courses in college. Subjects participated in small groups. They were presented with two versions of the following problem.

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 had been admitted to his two top choices: a small liberal arts college and an Ivy League university. 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 similar to his. The friends at the liberal arts college all 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 thought that he would go to the smaller college. However, he decided to visit both schools himself for a day.

He did not like what he saw at the private liberal arts college: Several people whom he met seemed cold and unpleasant; 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. Please say which school you think David should go to.

1. He should definitely go to the liberal arts college.
 2. He should probably go to the liberal arts college.
 3. It's a toss-up.
 4. He should probably go to the Ivy League university.
 5. He should definitely go to the Ivy League university.
- Please indicate why you made the recommendation you did.

The remaining subjects were presented with an identical problem except that the possibilities for error in David L.'s sample were highlighted by having him draw up a list of all the things that *might* be seen on the two campuses and then selecting randomly from among them for his actual schedule. The following was added to the second paragraph of the no-cue version.

He proceeded systematically to draw up a long list, for

both colleges, of all the classes which might interest him and all the places and activities on campus that he wanted to see. From each list, he randomly selected several classes and activities to visit, and several spots to look at (by blindly dropping a pencil on each list of alternatives and seeing where the point landed).

Open-ended responses to the probe question were coded (by a blind coder) as to whether subjects justified their choice by showing any recognition of the statistical questions involved—either the dubiousness of David L.'s impressions because of the quantity of his evidence or the superiority of his friends' testimony because of their greater experience. There was 90% agreement among coders as to the assignment of an answer to the statistical versus nonstatistical categories.

Results

When there was no sampling cue pointing to the probabilistic nature of David L.'s personal evidence base, 74% of the subjects recommended that David L. should go to the Ivy League university, which his friends didn't much like but where he enjoyed his day. When the sampling cue was present, this dropped to 56% ($\chi^2 = 5.38, p < .025$). Moreover, subjects in the probabilistic-cue condition were much more likely to refer to statistical considerations having to do with the adequacy of the sample. Fifty-six percent of probabilistic-cue subjects raised statistical questions in their open-ended answers, whereas only 35% of subjects in the no-cue condition did so ($p < .01$). Thus, when subjects are prompted to consider the possibilities for error that are inherent in a small sample of events, they are likely to shift to preference for large indirect samples over small personal ones, and their open-ended answers make it clear that it is statistical considerations that prompt this shift.

The findings of Study 3 are extremely ironic in that subjects are more likely to reject the superior personal evidence in the probabilistic-cue condition than to reject the inferior personal evidence in the control condition. This is because the same circumstances that serve to make the evidence superior in the probabilistic-cue condition also serve to make salient the extreme heterogeneity of the event population to be estimated and the small size of the personal sample of those events. It is important to note that when Study 3 is run with a "within" design, subjects assess the relative value of the personal evidence correctly; that is, they generally rate the quality of evidence

in the probabilistic-cue condition as superior to that in the control condition. In two slightly different within-design follow-ups to Study 3, subjects rated the probabilistic-cue sample as being superior to the sample in the control version. In one of the follow-ups (where subjects read the control problem and rated the quality of the personal evidence, then read the cue paragraph and compared the quality of the evidence there with the control version) four times as many subjects preferred the probabilistic-cue evidence as preferred the control evidence. In the other follow-up (where subjects actually acted as subjects in the control condition and then were shown the cue version), 40% more subjects preferred the probabilistic-cue evidence than preferred the control evidence.

Expertise and Statistical Reasoning

Study 4: Recognition of a Regression Effect in Sports and Acting

The studies we have just described indicate that subjects reason statistically when they recognize the heterogeneity of the events in question and the samplelike nature of their evidence about the events. If people are capable of learning from experience that events of a given kind are heterogenous and are produced in part by chance, then it should be possible to show that greater expertise in a domain is associated with a greater tendency to reason statistically in that domain. The two domains we selected to test this proposition were sports and acting. We anticipated that experience with sports would facilitate recognition of a regression effect in sports and that experience with acting would facilitate recognition of a regression effect in acting. Subjects were told about a small sample of extreme behavior followed by a larger sample of less extreme behavior. It was anticipated that inexpert subjects would generalize from the small sample and then would be obligated to give a causal explanation for the discrepancy between the small sample and the large sample. Expert subjects were expected to generalize less and to recognize that the discrepancy could be due to chance factors making the small sample appear extreme.

Method

Subjects were the same as those in Study 3. The problem presented to them was one of several in a study described as being aimed at finding out "how people go about explaining and predicting events under conditions of very limited information about the events." Subjects were given one of two nearly identically worded problems. One concerned a football coach who usually found that the most brilliant performers at tryout were not necessarily the best players during the football season, and the other concerned a repertory company director who usually found that the most brilliant performers at audition were not necessarily the best actors during the drama season. The full text of the football version is presented below.

Football. 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 younger 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?

The acting version of the problem was almost identical except that it was about Susan, the director of a student repertory company, who gets excited about "a young woman with great stage presence or a young man who gives a brilliant reading."

Subjects were asked which of the following explanations they preferred for the fact that the coach/director usually had to revise downward his or her opinion of the brilliant performers. The second alternative is the statistical one, suggesting that the explanation is simply that the tryout performances were uncharacteristically good for the "brilliant" performers.

1. Harold was probably mistaken in his initial opinion. In his eagerness to find new talent, he exaggerates the brilliance of the performances he sees at the try-out.
2. The brilliant performances at try-out are not typical of those boys' general abilities. They probably just made some plays at the try-out that were much better than usual for them.
3. The boys who did so well at try-out probably could coast through the season on their talent alone and don't put out the effort necessary to transform talent into consistently excellent performance.
4. The boys who did so well at try-out may find that their teammates are jealous. They may slack off so as not to arouse envy.
5. The boys who did so well at try-out are likely to be students with other interests. These interests would deflect them from putting all their energies into football.

Wording was altered very slightly for the acting version: "Boys" became "actors" and "try-out" became "audition."

Experience in sports was assessed by asking subjects whether they had played any organized team sports in high school or college. Those who had were defined as experienced. Experience in acting was defined as having had "more than a bit part" in a play in high school or college.

Results

It may be seen in Table 2 that experience affects the likelihood of preferring a statistical explanation for both the football version of the problem and the acting version. Most of the subjects with athletic team experience (a majority) preferred the statistical explanation for the football problem, whereas most of the subjects without team experience preferred one of the other, deterministic explanations. Most of the subjects with acting experience (a small minority) preferred the statistical explanation for the acting problem, whereas most of the subjects without acting experience preferred one of the deterministic explanations.

We do not wish to infer from these results that experience in a domain will make statistical explanations more salient for every kind of problem. Expertise brings a recognition of the causal factors at work in a domain as well as a recognition of the remaining uncertainty. When the problem can be approached with this expert causal knowledge, the expert may give answers that are less statistical, at least in form, than those of the novice. We may speculate that expertise reduces reliance on the representativeness heuristic, which encourages unreflective assumptions that the future will

resemble the past and that populations will resemble samples, and substitutes either statistical reasoning or reasoning in accordance with well-justified causal rules.

We should note also that it is possible that the tendency of experts to reason statistically may have less to do with knowledge of variability and uncertainty than with a subcultural norm for them to do so. The statistical answer may simply look more like a familiar, standard answer to the experts than to the nonexperts. For a correlational study such as Study 4, it is not easy to disentangle the undoubtedly related factors influencing statistical reasoning.

Domain Specificity of Inductive Rules

One of the major implications of the present viewpoint is that there should be a substantial degree of domain specificity of statistical reasoning. Its use should be rare for domains where (a) it is hard to discern the sample space and the sampling process, (b) the role of chance in producing events is unclear, and (c) no cultural prescription for statistical reasoning exists. We have noted that many of the studies showing people's failures in statistical reasoning examined judgments about events characterized by one or more of these factors. We have also noted that some of people's few demonstrated successes in statistical reasoning have been observed for people's reasoning about sports and academic achievements that seem to be characterized by clearer distributions for events, a more obvious role of chance,

Table 2
Percentage of Experienced and Inexperienced Subjects Who Preferred the Statistical Explanation for the Football and the Acting Problems

Problem	Subjects		χ^2	<i>p</i>
	Experienced	Inexperienced		
Football			3.10	.10
%	56	35		
<i>N</i>	52	26		
Acting			5.18	.025
%	59	29		
<i>N</i>	17	62		
Both versions			10.60	.001
%	57	31		
<i>N</i>	69	88		

and, probably, cultural prescriptions as well. These observations were made across studies, across tasks, and across subject populations, however.

To demonstrate that the same subjects dealing with the same tasks in the same experiment are more likely to reason statistically for some events characterized by uncertainty than for others, Jepson, Krantz, and Nisbett (in press) presented subjects with two broad classes of problems. The first class of problems dealt with events that are assessable by objective means, such as abilities, achievements, and physical illness. The second class dealt with events that are assessable only by subjective means, for example, personal preferences among objects, assessments of leadership potential, and judgments about the need for sexual fidelity in relationships. It was reasoned that, in general, it is relatively easy to apply statistical reasoning to objective events because one is likely to have some idea of their distributions (or to be able to guess what the distributions might look like because the units of measurement and the sample space are likely to be relatively clear). In general, also, the role of chance is likely to be relatively transparent for those objectively assessable events that in fact have been observed under repeated, relatively fixed conditions. Finally, cultural prescriptions to reason statistically probably exist for many such events. In contrast, none of these things is true for most events that can be assessed only by subjective means. The first problem below is an example of the Jepson et al. *objective* problems; the second is an example of the *subjective* problems.

Championship Selection Problem

Two sports fans are arguing over which sport—baseball or football—has the best (most accurate) playoff 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?

Class Selection Problem

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?

Subjects wrote open-ended answers to problems of each type. These were coded as to whether they reflected the use of statistical principles—chiefly the law of large numbers or the regression principle—or not. An example of a nonstatistical answer for the championship problem is the following: “Super Bowl, because of neutral ground and also a one shot deal. Either you make it or break it—one chance. The pressure is on to perform the team's best.” An example of a statistical answer is the following: “World Series is better. Anyone can get lucky for one game, but it is harder to be lucky for four. Besides, being home or away is part of the game, you don't play on neutral ground during the season.” An example of a nonstatistical answer for the class-choice problem is the following: “He's got to choose for himself.” An example of a statistical answer is: “You can't tell from one time—thus a survey that is over a longer range is better. Although Henry's idea of a good class could be different from most students.”

Statistical answers were much more common for problems about objective events than for problems about subjective events. Forty-one percent of the answers for the former were statistical; the range of mean percentage statistical answers across problems was 30%–93%. Only 12% of the answers to subjective problems were statistical; the range was 5%–16%.

The results also showed that subjects were consistent in their tendency to give statistical answers. Those who gave statistical answers for any given problem were more likely to give them for any other. This tendency was correlated with both verbal and mathematical scores on the Scholastic Aptitude Test.

The results of the Jepson et al. (in press) study show that the same subjects in the same context answering the same general kind of problem are more likely to give statistical answers for a set of problems dealing with abilities than for a set dealing with more subjective attributes. The results do not rule out the possibility that it is problem structure or the exact nature of the required statistical rule that produces the difference among problems rather

than their content. To control for this, Fong, Krantz, and Nisbett (Note 3) developed five separate problem structures for which somewhat different versions of the law of large numbers were required to produce the most appropriate statistical solution. For example, in one structure a small personal sample of events had implications that were different from those of a large sample collected by someone else (as in the class-choice problem above). Subjects were asked to indicate which sample was the better guide to action and why. Another structure presented a simple regression problem in which subjects were asked to explain why an extreme outcome for a small sample was not maintained in a larger sample. For each problem structure, two or more problem versions were constructed, some dealing with objective events and others with subjective ones. With structure and required rule type thus controlled, there was still a marked difference in the percentage of statistical answers for the two kinds of events.

It should be noted that an alternative explanation still exists even for the Fong et al. (Note 3) results. This is the possibility that statistical answers are not as appropriate for the subjective problems as for the objective problems. This objection is not readily answerable with our present level of knowledge about the uncertainty of events. We can only urge our view that statistical answers were fully appropriate for all problems because all problems involved high degrees of uncertainty. But it must be acknowledged that other people having the relevant statistical training might not agree.

The Effects of Training in Statistics on Reasoning About Everyday Events

Perhaps the most important implication of the present view is that statistical reasoning about everyday events should be highly trainable. A major reason for optimism is that, as we have just demonstrated, people's intuitive reasoning skills include strategies that may be called statistical heuristics. Formal training in statistics, therefore, should represent less a grafting on of procedures than a refinement of preexisting ones. Formal training seems likely to improve reasoning for three distinct, but mutually supportive, reasons.

1. *Training in statistics is apt to facilitate the recognition of event distributions and their statistical parameters.* It can be difficult to apply rules such as the law of large numbers unless the units of evidence can be identified and, hence, the sample space and distribution of the event units. It seems likely that training in statistics could provide quite general skills in construing evidence in such a way that it can be properly unitized, the sample space identified, and parameters recognized. Training in probability theory, especially in permutations and combinations, should be particularly likely to be helpful in this regard as should test theory, which requires the student to recognize, for example, the different reliability of tests composed of different numbers of units—items, trials, occasions, and so on. But ordinary inferential statistics also introduces a fair amount of instruction in unitizing evidence: In order to measure the corn yield of a type of seed, for example, it is necessary to measure the yield for some unit of sampling (individual plot, individual farm, etc.). Statistics courses should also make it easier for people to think usefully about parameters of distributions—about central tendencies and about dispersion.

2. *Training in statistics is apt to facilitate the recognition of the role of chance in producing events.* A major concept of parametric statistics is that of error. Every inferential test features an estimate of error, and much of statistical training centers on questions of allocation of effects to the systematic category versus the random category. It seems possible that the focus on the concept of error in statistics might heighten the salience of uncertainty in daily life. A second major concept that might be helpful in recognizing uncertainty is the formal notion of sampling. This might be of general use in construing evidence as a (possibly small and inaccurate) sample from a (possibly heterogeneous) population.

3. *Training in statistics is apt to improve the clarity and accessibility of statistical rules and should expand the repertoire of statistical rules.* In effect, statistics training should hone intuitive heuristics into more precise tools. It seems clear that without training some statistical rules are poorly understood at any level of abstraction and in any context. Rules of covariation assessment and some versions of the regression principle are particularly dif-

ficult and may not even be represented in most people's intuitive repertoire. But even relatively intuitive rules such as the law of large numbers have nearly limitless corollaries and implications, some of which may be much easier to discern with formal training. Although people understand the "law of small large numbers," they may not be able to extend the principle to numbers of nonexperiential magnitudes without formal training. It may be the lack of a fully formal understanding of the law of large numbers that prevents people from applying it in the maternity ward problem, for example, where the sample sizes involved (15 and 45) are not often represented in people's everyday experience of sampling and variability.

The evidence indicates that statistical training does indeed have profound effects on people's reasoning about everyday life events. In one series of studies, Krantz, Fong, and Nisbett (Note 4) examined four groups of subjects differing widely in educational level. Subjects, who were college students with or without statistical training, graduate students with a fair amount of statistical training, or PhD level scientists with several years of training, were presented with one of a pair of restaurant problems. In each problem, a protagonist experienced a truly outstanding meal on the first visit to a restaurant but was disappointed on a repeat visit. The subjects were asked to explain, in writing, why this might have happened. A subject's explanation was classified as nonstatistical if it assumed that the initial good experience was a reliable indicator that the restaurant was truly outstanding and attributed the later disappointment to a definite cause, such as a permanent or temporary change in the restaurant (e.g., "Maybe the chef quit") or a change in the protagonist's expectation or mood. The explanation was classified as statistical if it suggested that meal quality on any single visit might not be a reliable indicator of the restaurant's overall quality (e.g., "Very few restaurants have only excellent meals; odds are she was just lucky the first time"). Statistical explanations were coded as to how articulate they were in indicating that a single visit may be regarded as a small sample and, hence, as unreliable. Explanations were thus coded as falling into one of three categories: (1) nonstatistical, (2) poor statistical,

and (3) good statistical. The frequencies in each of these categories were used to define two dependent variables: *frequency* of statistical answers, defined as the proportion of responses in Categories 2 and 3, and *quality* of statistical answers, defined as the proportion of Category 2 and 3 answers that were Category 3.

The two versions of the restaurant problem differed. The probabilistic-cue version included a random mechanism for selection from the menu: The protagonist did not know how to read a Japanese menu and selected a meal by blindly dropping a pencil on the menu and observing where the point lay. The other version had no such cue. Within each group tested, half of the subjects received the cue and half did not.

The effects of training on both dependent measures were dramatic. College students without statistical training almost never gave an answer that was at all statistical unless the problem contained the probabilistic cue, in which case about half of the answers were statistical. In contrast, more than 80% of the answers of PhD-level scientists were statistical, whether or not there was a cue. Quality of statistical answers also depended on level of training. Only 10% of the statistical answers by untrained college students were rated as good, whereas almost 80% of the statistical answers by PhD-level scientists were rated as good. It is interesting that although the presence of the probabilistic cue was very important in determining whether less trained subjects would give a statistical answer, it did not affect the quality of statistical answers for subjects at any level of training. Apparently probabilistic cues can trigger the use of statistical heuristics, but they do not necessarily improve the quality of answers: the appropriate skills must be in the individual's repertoire to insure good quality.

The preceding study confounds training and native mathematical ability, but subsequent studies both avoid the confounding with ability and show that statistical training influences inductive reasoning outside the classroom and laboratory. Krantz et al. (Note 4) conducted a telephone "survey of opinions about sports." Subjects were males who were enrolled in an introductory statistics course and who admitted to being at least somewhat knowl-

edgeable about sports. One hundred subjects were randomly selected and surveyed during the first 2 weeks of the term they were enrolled in statistics. Another 93 students were surveyed at or near the end of the term. In addition to filler questions on NCAA rules and NBA salaries, subjects were asked questions for which a statistical approach was relevant, as in the example below.

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?

Most subjects answered this question in a nonstatistical way, invoking notions like "too much press attention" and "slacking off." Some subjects answered the question statistically (e.g., "There are bound to be some rookies who have an exceptional season; it may not be due to any great talent advantage that one guy has over the others—he just got a good year").

The statistics course increased the percentage of statistical answers and also increased the quality of statistical answers. The course also markedly influenced both the frequency and the quality of statistical answers to another question asking subjects to explain why .450 batting averages are common the first 2 weeks of the baseball season but are unheard of as a season average. In all, the course had a significant effect on three of the five questions asked.

Finally, Fong et al. (Note 3) showed that even a very brief training procedure can suffice to affect markedly subjects' answers to problems about everyday events. There were two major elements in their training package: One covered formal aspects of sampling and the law of large numbers and the other showed how to use sampling notions as a heuristic device in modeling problems. In the sampling instruction, subjects received definitions of population and sample distributions, a statement of the law of large numbers, and a demonstration (by drawing colored gumballs from a glass vase) that a population distribution is estimated more accurately, on the average,

from larger samples. The modeling, or mapping, instruction consisted of three problems (in the general style of the restaurant problem and similar to the subsequent test problems), each followed by a written solution that used the law of large numbers and emphasized the analogy between amount of evidence and size of sample.

There were four major conditions: a control group given no instruction and three experimental groups—one given sampling training only, one given mapping training only, and one given both types of training. The subjects were adults and high school students. The test consisted of 15 problems. Five of these had clear probabilistic cues, five dealt with objective attributes such as abilities or achievements, and five dealt with subjective judgments. Training effects were marked for all three problem types, for both the frequency of statistical answers and the quality of statistical answers. Sampling training and mapping training were about equally effective, and in combination they were substantially more effective than either was alone. A particularly encouraging finding is that training showed no domain specificity effects. In a companion study, Fong et al. (Note 3) showed that it made no difference to performance whether mapping training had been on probabilistic-cue problems, objective attribute problems, or subjective judgment problems. The latter finding suggests that training on specific problem types can be readily abstracted to a degree sufficient for use on widely different problem types.

The work on training should not be taken to indicate that a statistical education is sufficient to guarantee that people will avoid errors in inductive reasoning. Kahneman and Tversky (1982; Tversky & Kahneman, 1971, 1983) have shown repeatedly that statistical expertise provides no such guarantee against errors. On the other hand, it should also be noted that courses in statistics do not emphasize ways to use statistical principles in everyday life. Were they to do so, one might see much larger differences between the educated and the uneducated than we have found.

Normative Considerations

People can apply statistical reasoning to a wide range of problems of an everyday life

sort. The use of statistical reasoning seems to be increased by greater clarity about the sample space and the sampling process, by recognition of the role played by chance, and by cultural prescriptions to apply statistical reasoning. As a consequence, statistical reasoning appears to be more prevalent for events that are assessed by objective means than for events that are assessed only subjectively. People apply statistical reasoning more frequently and more aptly after formal training in statistics.

Exactly what is the content of lay statistical heuristics? What is the range of problems to which they can be applied? Which statistical principles are well entrenched in the repertoire of formally untutored people and which are not represented or even counterintuitive? Is there an improvement or a worsening of statistical reasoning as one moves from laboratory studies of reasoning to studies of *in vivo* reasoning about analogous problems in their appropriate ecological context? Exactly how much improvement in people's ability to reason statistically could be expected from traditional formal education in statistics? How could educational practices be improved so as to amplify the real world consequences of training?

We refrain from speculating about these questions, despite their importance, because they are fundamentally empirical in nature. It would be more appropriate for us to sketch the normative implications of the work to date rather than to try to be prescient about future matters of fact. We have addressed many of the normative implications of the previous work on limits elsewhere (Krantz, 1981; Nisbett, 1981; Nisbett, Krantz, Jepson, & Fong, 1982; Stich & Nisbett, 1980; Thagard & Nisbett, 1983), and we have addressed at some length some of the normative implications of our work on individual differences (Jepson, Krantz, & Nisbett, in press), but it would be useful to summarize the general normative implications of the work presented in this review.

Ecological Representativeness of Problems Showing Errors

One criticism of the literature showing errors in inductive reasoning has been to argue that they are the result simply of examining

people's judgment about particular kinds of problems, and in a particular kind of context, where judgments are particularly likely to be fallible (e.g., Cohen, 1981; Dennett, 1981; Lopes, 1982). The tenor of this criticism is that the studies show more about the cleverness of experimenters than they do about the real world failures of lay people.

The accusation that psychologists have been devising parlor tricks, which people are susceptible to in the laboratory context but either do not encounter or could solve in real world contexts, seems less plausible in view of the research reported here. First, for each problem we have reported, some of the subjects showed by their answers (and often by the rationales for their answers, subsequently elicited) an appreciation of the statistical principles that in previous work other subjects failed to appreciate. It seems more reasonable to explain the success of some of our subjects by saying that they are more skilled at statistical reasoning than the other subjects rather than to explain it by saying that they saw through the experimenters' tricks. Second, the factors that make statistical reasoning more or less likely, for example, recognition of heterogeneity and of the role played by chance, do not sound like factors that make people more or less dupable by experimenters but rather like factors that make the appropriateness of statistical reasoning more or less obvious. Third, statistical training markedly influences answers to the sort of problems we studied. This suggests that it is not problem- or context-produced illusions that make people unable to solve statistical problems but simply lack of statistical knowledge.

"Satisficing" in Decision Making and Inductive Reasoning

Since Simon's (1957) important work on decision making, it has been a standard part of normative analysis to point out that, because of time pressures and other constraints, it may be quite sensible for people to depart from formal decision models. This corner-cutting practice is called *satisficing* (in distinction to the presumed *optimizing* that would result from the formal procedures). This same defense is often applied to people's failures to reason statistically (Einhorn & Hogarth, 1981;

Miller & Cantor, 1982; Nisbett & Ross, 1980). People who study inductive reasoning seem to have presumed that normatively correct inductive reasoning is usually more laborious and time-consuming than is purely intuitive inductive reasoning, just as formal decision making is usually more time consuming than is intuitive decision making. The present work makes it clear that this presumption cannot be imported uncritically into the realm of inductive reasoning. Exclusively causal reasoning and the search for values on putatively relevant causal factors can be extraordinarily laborious. Statistical reasoning, once it is mastered, can be very rapid, even automatic. We found it striking, for example, to contrast the answers given over the telephone by subjects in the sports survey by Krantz et al. (Note 4). Some subjects doggedly persisted in causal explanations for problems such as the rookie of the year, sophomore slump ("Well, the success goes to their heads . . . and there's pressure to keep up the performance after the great first year . . . and . . ."), which did not seem to satisfy even the subjects who were generating them. Other subjects generated quick, crisp statistical explanations ("They just happened to have a first year that was better than their lifetime average") that still left them free to explore possible causal explanations.

It is also important to note that one has very little sense of subjects' *choosing* inferential strategies when one reads such protocols, in the sense that people may choose a formal decision-making strategy over an intuitive one. Rather, subjects either seem to spontaneously pursue a statistical approach or they do not. Formal decision-making procedures involve novel and counterintuitive practices such as drawing tree diagrams and multiplying probability and utility assessments. On the other hand, many statistical procedures, as we have shown, have their simpler intuitive counterparts in the equipment of everyday thought. There is no reason to presume that these will be any more cumbersome to use or will require any more conscious deliberation to access than will other intuitive approaches. Similarly, training in formal statistics may produce automatic, nonreflective transfer to everyday problems in a way that formal decision training would not.

Evolution and Inductive Reasoning

Many people have responded to the work demonstrating inferential errors by assuming that the errors are either exaggerated or that they are the incidental by-product of some overwhelmingly useful inferential procedure that happens to go astray under ecologically rare circumstances. This is essentially the argument from design, and several philosophers have endorsed it. (See, e.g., Cohen, 1979; Dennett, 1978, 1981, Note 1; Lycan, 1981. Einhorn & Hogarth, 1981, have presented several very compelling arguments against the design view, and we shall not repeat them here.)

Endorsement of the evolutionary, or design, view requires a rather static, wired-in assumption about the nature of inferential procedures. Philosophers are not alone in making this assumption, it should be noted. Psychologists who are wont to presume unlimited plasticity in social behavior often seem to presume complete rigidity in inferential rules, as if these could be influenced at most by maturation. In our view, there are few grounds for such a presumption. Whatever may be true for deduction, there are good grounds for assuming that inductive procedures can be changed. Renaissance physicians adhered to the doctrine of signatures, an inductive system with both descriptive and procedural components. Modern physicians have curtailed the scope of the representativeness heuristic in their daily inferential lives. (Although it is still relied on in interviews of applicants to medical schools. See Dawes, 1980, for a description of the doctrine of signatures at work in the admissions process for American universities.) Sophisticated causal analysis and statistical reasoning will eventually result in the further curtailment of simple intuitive heuristics, for physicians and for everyone else.

Although we see no merit in an evolutionary defense of the inferential behaviors that happen to characterize American college students in the latter part of the 20th century, we see a powerful argument in the work we have reviewed for the role of cultural evolution. It does not require unusual optimism to speculate that we are on the threshold of a profound change in the way that people reason inductively. The range of events that scientists can

think about statistically has been increasing slowly but in a decided, positively accelerated fashion at least since the 17th century. The work of Kahneman and Tversky may be regarded as the most recent and one of the most dramatic inflection points on that curve.

We believe that, with a lag in time, lay people have been following a similar curve of ever-widening application of statistical reasoning. Most people today appreciate entirely statistical accounts of sports events, accident rates, and the weather; also, we found many subjects who gave statistical explanations even for subjective events such as disappointment about meals served by a restaurant; and Piaget's young subjects reasoned about the behavior of randomizing devices with a sophistication that seems quite unlikely for people of earlier centuries. Will our own descendants differ as much from us as we do from Bernoulli's contemporaries?

Reference Notes

1. Dennett, D. C. *True believers: The intentional strategy and why it works*. Unpublished manuscript, Tufts University, 1983.
2. Kahneman, D. Personal communication, 1982.
3. Fong, G. T., Krantz, D. H., & Nisbett, R. E. *Improving inductive reasoning through statistical training*. Unpublished manuscript, University of Michigan, 1983.
4. Krantz, D. H., Fong, G. T., & Nisbett, R. E. *Formal training improves the application of statistical heuristics to everyday problems*. Unpublished manuscript, Bell Laboratories, Murray Hill, New Jersey, 1983.

References

- Borgida, E., & Nisbett, R. E. The differential impact of abstract vs. concrete information on decisions. *Journal of Applied Social Psychology*, 1977, 7, 258-271.
- Cohen, L. J. On the psychology of prediction: Whose is the fallacy? *Cognition*, 1979, 7, 385-407.
- Cohen, L. J. Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, 1981, 4, 317-331.
- Dawes, R. M. You can't systemize human judgment: Dyslexia. *New Directions for Methodology of Social and Behavioral Science*, 1980, 4, 67-78.
- Dennett, D. C. *Brainstorms*. Montgomery, Vt.: Bradford Books, 1978.
- Dennett, D. C. Three kinds of intentional psychology. In R. Healey (Ed.), *Reduction, time and reality*. Cambridge, England: Cambridge University Press, 1981.
- Einhorn, H. J., & Hogarth, R. M. Confidence in judgment: Persistence of the illusion of validity. *Psychological Review*, 1978, 87, 395-416.
- Einhorn, H., & Hogarth, R. M. Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, 1981, 32, 53-88.
- Hacking, I. *The emergence of probability*. New York: Cambridge University Press, 1975.
- Hamer, C. *"Good-fishin'!"* (2nd ed.). St. Paul, Minn.: P.F., Inc., 1981.
- Hogarth, R. M. *Judgement and choice*. New York: Wiley, 1980.
- Hume, D. *An inquiry concerning human understanding*. Indianapolis, Ind.: Bobbs-Merrill, 1955. (Originally published, 1748.)
- Jepson, C., Krantz, D. H., & Nisbett, R. E. Inductive reasoning: Competence or skill? *Behavioral and Brain Sciences*, in press.
- Kahneman, D., Slovic, P., & Tversky, A. *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press, 1982.
- Kahneman, D., & Tversky, A. Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 1972, 3, 430-454.
- Kahneman, D., & Tversky, A. On the psychology of prediction. *Psychological Review*, 1973, 80, 237-251.
- Kahneman, D., & Tversky, A. On the study of statistical intuitions. *Cognition*, 1982, 11, 237-251.
- Krantz, D. H. Improvements in human reasoning and an error in L. J. Cohen's. *Behavioral and Brain Sciences*, 1981, 4, 340-341.
- Law of averages a reality for ballplayers. *Los Angeles Times*. June 7, 1981, Part 3, p. 6.
- Lopes, L. L. Doing the impossible: A note on induction and the experience of randomness. *Journal of Experimental Psychology*, 1982, 8, 626-636.
- Lycan, W. G. "Is" and "ought" in cognitive science. *Behavioral and Brain Sciences*, 1981, 4, 344-345.
- Mill, J. S. *A system of logic ratiocinative and inductive*. Toronto, Ontario, Canada: University of Toronto Press, 1974. (Originally published, 1843.)
- Miller, G. A., & Cantor, N. Book review of Nisbett, R., & Ross, L. *Human inference: Strategies and shortcomings of social judgment*. *Social Cognition*, 1982, 1, 83-93.
- Nisbett, R. E. Lay arbitration of rules of inference. *Behavioral and Brain Sciences*, 1981, 4, 349-350.
- Nisbett, R. E., & Borgida, E. Attribution and the psychology of prediction. *Journal of Personality and Social Psychology*, 1975, 32, 932-943.
- Nisbett, R. E., Krantz, D. H., Jepson, C., & Fong, G. T. Improving inductive inference. In D. Kahneman, P. Slovic, & A. Tversky (Eds.) *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press, 1982.
- Nisbett, R. E., & Ross, L. *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, N.J.: Prentice-Hall, 1980.
- Piaget, J., & Inhelder, B. *The origin of the idea of chance in children*. New York: Norton, 1975. (Originally published in 1951.)
- Quattrone, G. A., & Jones, E. E. The perception of variability within in-groups and out-groups: Implications for the law of large numbers. *Journal of Personality and Social Psychology*, 1980, 38, 141-152.

- Silka, L. Effects of limited recall of variability on intuitive judgments of change. *Journal of Personality and Social Psychology*, 1981, 40, 1010-1016.
- Simon, H. A. *Models of man: Social and rational*. New York: Wiley, 1957.
- Stich, S., & Nisbett, R. E. Justification and the psychology of human reasoning. *Philosophy of Science*, 1980, 47(2), 188-202.
- Thagard, P., & Nisbett, R. E. Variability and confirmation. *Philosophical Studies*, 1982, 42, 379-394.
- Thagard, P., & Nisbett, R. E. Rationality and charity. *Philosophy of Science*, 1983, 50, 250-267.
- Tversky, A., & Kahneman, D. Belief in the law of small numbers. *Psychological Bulletin*, 1971, 76, 105-110.
- Tversky, A., & Kahneman, D. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 1973, 5, 207-232.
- Tversky, A., & Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science*, 1974, 185, 1124-1131.
- Tversky, A., & Kahneman, D. Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 1983, 90, 293-315.
- Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. M. Perceiving the causes of success and failure. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, N.J.: General Learning Press, 1972.

Received April 13, 1983

Revision received July 9, 1983 ■