

Insensitivity to Sample Bias: Generalizing From Atypical Cases

Ruth Hamill
University of Michigan

Timothy DeCamp Wilson
University of Virginia

Richard E. Nisbett
University of Michigan

Two experiments were conducted to determine whether subjects take into account the representativeness of a sample before generalizing from the sample to a population. Subjects were presented with vivid one-case samples of populations—a welfare recipient in one study and a prison guard in another. Subjects were then asked to rate the population (of welfare recipients or prison guards) on a number of dimensions. Exposure to the sample case influenced attitudes about the population whether subjects were told nothing about the typicality of the case, were told that the case was highly typical of the population, or were told that the case was highly atypical of the population. The results suggest that, at least when information about sample bias is pallid and information about the nature of the sample is vivid, people may make unwarranted generalizations from samples to populations.

Much of our knowledge of the world may be thought of as inductive generalizations from samples to populations. For example, beliefs about members of a particular occupational group or ethnic group often are based on inductive generalizations from personal samples of the occupational or ethnic population. Similarly, beliefs about a particular city often are influenced by one's personal sample of the population of experiences to be had in the city, and one's beliefs about a friend are based largely on samples of the population of the friend's behaviors.

For inductive inferences to qualify as valid, they should be based on samples of reasonable size and representativeness. There is substantial evidence, however, that people are not very sensitive to sample size considerations. Tversky and Kahneman (1971; Kahneman & Tversky, 1972) have shown that people (including even scientists with strong backgrounds in statistics) are insufficiently impressed by large samples and are unduly responsive to small samples. In daily life, this tendency could result in judgments and actions that are not well justified by the available evidence. For example, people may buy a car or take a college course (Borgida & Nisbett, 1977) on the recommendation of one or two people and may not trouble themselves to seek out larger samples even when additional information is readily available.

This research was supported by Grants BNS 75-23191 and BNS 79-14094 from the National Science Foundation. The authors are indebted to Ronald Lemley and Harold Neighbors for their portrayals of prison guards, to Amos Tversky for design suggestions, and to Lee Ross and Keith Sentic for helpful comments on a previous draft of the manuscript.

Ruth Hamill is now at the Department of Political Science, State University of New York at Stony Brook.

Requests for reprints should be sent to Richard E. Nisbett, Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48109.

The literature also provides some indications that people may be insufficiently sensitive to the possibility of sample bias. Nisbett and Borgida (1975) presented descriptions of two psychology experiments to subjects and asked them to predict the behavior of the population of participants in the experiments.

Prior to making their predictions, some subjects were shown videotaped interviews with a sample of two participants, both of whom had behaved in an extreme, unanticipated way. Subjects exposed to the sample predicted that the extreme behavior was the modal behavior for the population. They did so to the same extent regardless of whether the basis of selection of the samples was unspecified (and therefore might have been biased in some way) or was explicitly described as random.

A study by Ross, Amabile, and Steinmetz (1977) also suggested that people may be insensitive to sample bias. They asked two college students to participate in a general knowledge quiz. One subject was designated as the questioner and was asked to generate 10 "challenging but not impossible" questions from his store of general knowledge. The other subject, the contestant, attempted to answer these questions and then rated both his own level of general knowledge and the questioner's. Contestants rated the questioners as being much more knowledgeable than themselves. They were apparently insufficiently responsive to the obvious source of bias in the questions generated by the questioners—that is, those questions were drawn from that small portion of general knowledge for which the questioner happened to know all the answers.

The Nisbett and Borgida (1975) study suggests that people may be insufficiently sensitive to the superiority of random samples over samples for which the basis of selection is unspecified. The Ross et al. (1977) study suggests that some sources of extreme bias may not be highly salient to people or that they may make insufficient adjustments for the bias. These findings raise the possibility that people will make inductive generalizations even when it is clear that the sample is highly biased with respect to relevant parameters.

Nisbett and Borgida (1975) and Nisbett and Ross (1980) speculated that an important reason for people's imperfect utilization of inductive principles is that information about the sample itself, such as concrete details about a particular person, is vivid and interesting. Information pertinent to the kinds

of inferences that can be drawn from the sample, such as the size of the sample or how biased it might be, is generally pallid and uninteresting. If, as Nisbett and his colleagues propose, information is utilized in inference in proportion to its vividness, people exposed to a biased but highly vivid sample might generalize to the population even when they are informed that the sample is highly atypical of the population in some important respect.

Study 1

Method

Overview

Attitudes toward the population of welfare recipients in the U.S. were assessed. Subjects were randomly assigned to one of six conditions—four experimental and two control. All experimental subjects read a booklet containing a description of an irresponsible woman who had been on welfare for many years. Experimental subjects in the *typical sample* condition read statistics about welfare recipients that made it clear that the woman was typical of welfare recipients with respect to the length of her stay on welfare. Subjects in the *atypical sample* condition read statistics that made it clear that the woman had been on welfare much longer than was common. Half of the subjects within each of these experimental conditions were provided with the sampling information *before* they read the article, and half were presented with the information *after* they read the article, thus forming four experimental groups. After reading the description of the welfare case, all experimental subjects responded to a dependent measure questionnaire on attitudes toward welfare recipients.

Two control groups also participated. None of the subjects in either control group read the description of the welfare recipient, and all responded to the dependent measure questionnaire. To determine whether the (quite favorable) statistical information presented to atypical sample experimental subjects would by itself make subjects more favorably disposed toward welfare recipients, one group of control subjects, the *informed* group, was presented with this statistical information. The *uninformed* control group received no information prior to answering the questionnaire. Some of these subjects, however, were given a short quiz concerning their knowledge about welfare recipients. This allowed us to ascertain naive subject assumptions about length of stay on welfare. After completing the dependent measure questionnaire, all subjects were debriefed.

Subjects

The subjects were 127 University of Michigan introductory psychology students of both sexes.¹ Subjects were randomly assigned to one of the six conditions as they arrived at the experiment. They were seated in individual booths and given the appropriate instructions and materials for their condition.

Experimental Conditions

Subjects in the two experimental conditions were presented with a vivid magazine article describing a welfare case. For these subjects, the study was presented as a survey dealing with college student tastes in magazine articles. They were told they would read an article from a popular magazine and then express their opinions about its content and style. The article (Sheehan, 1975) was an abridged version of a piece from the *New Yorker* and presented the "sample" case for subjects. The article provided a detailed description of the history and current life situation of a 43-year-old, obese, friendly, irresponsible, ne'er-do-well woman who had lived in New York City for 16 years, the last 13 of which had been spent on welfare. The woman had emigrated from Puerto Rico after a brief, unhappy teenage marriage that produced three children. Her life in New York was an endless succession of common-law husbands, children at roughly 18-month intervals, and dependence on welfare. She and her family lived from day to day, eating high-priced cuts of meat and playing the numbers on the days immediately after the welfare check arrived, and eating beans and borrowing money on the days preceding its arrival. Her dwelling was a decaying, malodorous apartment overrun with cockroaches and filled with shoddy, plastic-covered furniture bought on time at outrageous prices. Her children attended school as they pleased. They began to run afoul of the law in their early teens, and by early adulthood they were hopelessly enmeshed in a life of drugs, numbers-running, and welfare.

Our sample welfare case is a vivid embodiment of cultural stereotypes about welfare recipients, but she is in fact quite atypical. Of the women aged 18-54 in the United States who have been on welfare at all in a 10-year period, only a minority have been on welfare for more than 2 consecutive years or for more than 4 total years in any 10-year period. The proportion of all welfare recipients for whom more than half of total financial support comes from welfare for as much as 9 of the 10 years, is quite small—less than 10% (Rein & Rainwater, 1977). Thus our sample welfare case was actually quite uncharacteristic of the population of all welfare recipients with respect to the duration and extent of her dependence on welfare. A version of this statistical information was provided to subjects in the atypical sample condition.

Atypical sample condition. For two groups of subjects, the atypicality of our sample case was

clearly specified. The article was preceded (for one group of 20 subjects) or followed (for another 20 subjects) by a prominent "Editor's Note," stating the following statistics:

Statistics from the New York State Department of Welfare show that the average length of time on welfare for recipients between the ages of 40 and 55 is 2 years. Furthermore, 90% of these people are off the welfare rolls by the end of 4 years. The author of the following (preceding) article interviewed a woman in the age range described by these statistics. The account below (above) is based on these interviews.

Typical sample condition. For the subjects in this condition, the Editor's Note (which preceded the article for 20 subjects and followed the article for the other 20) was written so as to make it seem as though our sample case was a very typical welfare recipient. The average length of stay on welfare was described as 15 years, and subjects were told that "90% of these people are still on the welfare rolls after 8 years."

In all experimental conditions, the dependent measures assessing attitudes toward welfare were introduced by stating that

It is often the case with articles on political and social issues that people's general views on these topics affect their reactions to the articles. In order to determine whether this is the case with the article you have just read, we would appreciate it if you would answer the questions below.

After responding to the dependent measures, subjects were asked to recall as best they could the statistics presented to them about the average stay on welfare.

Control Conditions

Control subjects were assigned to one of two groups, although all control subjects were told that their attitudes toward welfare would be assessed, and none read about the welfare case before responding to the dependent measures.

Informed condition. One group of subjects ($n = 21$) was asked to respond to a "Did you know?" quiz about welfare before completing the dependent measure questionnaire. In addition to several filler items, subjects were asked to indicate whether or not they had previously known that "the average length of time on welfare for recipients between the ages of 40 and 55 in New York state is 2 years" and that "90% of the people in the above age range in New York are off the welfare rolls by the end of 4 years." This statistical information

¹There were no sex differences for any of the major dependent variables in either Study 1 or Study 2.

was identical to that presented to atypical sample subjects, and it was expected to provide information about length of stay on welfare that would be much more favorable than subjects' previously-held beliefs. This manipulation made it possible to assess the effects on attitudes of the favorable statistical information alone, without exposure to the sample welfare case.

Uninformed condition. Subjects in the uninformed condition ($n = 26$) were given no statistical information and simply responded to the dependent measures.

Assessing prior beliefs about welfare. Some of the subjects in the uninformed control condition ($n = 16$) were given a quiz about welfare prior to the dependent measures. Among several filler questions, subjects were asked, "What is the average length of time on welfare for recipients between the ages of 40 and 55 in New York State?" and "After how many years are 90% of the recipients in the above age range off welfare?" This questionnaire allowed us to determine naive beliefs about the average length of stay on welfare.

Dependent Measures

Subjects responded to a number of filler items, including several items assessing attitudes toward the welfare system.² Subjects then responded to seven 5-7-point scales assessing their attitudes toward welfare recipients. Two examples are given below.

Even if they had ample financial support, many people on welfare would lead disorganized, socially unproductive lives (1 = strongly agree, 6 = strongly disagree).

How hard do people on welfare work to improve their situations? (1 = not at all hard, 5 = extremely hard).

Responses to the seven items were summed to create the dependent measure.

Results

Manipulation Checks

At the end of the experiment, subjects in the experimental conditions were asked to recall the statistics about welfare recipients that had comprised the sampling information manipulation. Subjects in the atypical sample condition had been told that the average length of stay on welfare for New York State recipients between the ages of 40 and 55 was 2 years and that 90% were off the rolls by the end of four years. Subjects' recall was fairly accurate: Their mean estimates were that the average stay was 2.85 years and that 90%

were off the rolls by 4.25 years. Subjects in the typical sample condition were told that the average length of stay was 15 years and that 90% of recipients were still on the rolls at the end of 8 years. Subjects in this condition were also fairly accurate: Their mean estimates were that the average stay was 14.8 years and that 90% were still on the rolls at the end of 11.2 years. Subjects who received the statistical information before reading the article did not differ significantly in the accuracy of their recall from subjects who received the information after reading the article (and therefore much closer to the time of the recall measure).

Effects of Exposure to the Sample

Table 1 presents mean attitudes toward welfare recipients for the experimental conditions and for the combined controls.³ The manipulation of order of presentation of sampling information (before or after reading the article) had only trivial and nonsignificant effects on attitudes toward welfare recipients for both typical and atypical sample conditions, so results were collapsed across this variable for both conditions. Similarly, the two control groups differed only trivially, so results for the two groups were combined.

It may be seen that exposure to the sample welfare case produced unfavorable attitudes toward welfare recipients. Results from Dunnett's procedure for comparing each experimental group mean with the control group mean show that there was a significant difference between the typical sampling information group and the control group, $t(121) = 2.96$, $p < .01$, as well as a significant difference between the atypical sampling information group and the control group, $t(121) = 2.50$, $p < .05$. Moreover, the typical sample group and the atypical sample group differed only trivially from each other: The uncor-

² The manipulations had no effect on attitudes toward the welfare system.

³ Two subjects from the atypical sample information group and one from the typical sample information group failed to answer at least one of the questions concerning welfare recipients and were therefore excluded from the analysis of attitudes toward recipients.

Table 1
*Means and Standard Deviations of Attitudes
 Toward Welfare Recipients for Groups
 Exposed to a Typical Sample, an Atypical
 Sample, or No Sample*

Measure	Subject group		
	Typical sample (<i>n</i> = 39)	Atypical sample (<i>n</i> = 38)	Combined control condition (<i>n</i> = 47)
<i>M</i>	20.18 ^a	20.68 ^b	23.45
<i>SD</i>	5.04	4.80	5.36

Note. The higher the mean, the more favorable the attitude toward welfare recipients. Univariate $F(2, 121) = 5.22, p < .0007$.

^a Differs from the control condition at $p < .01$, based on Dunnett's procedure.

^b Differs from the control condition at $p < .05$, based on Dunnett's procedure.

rected (and therefore overly liberal) t contrasting the two groups was .44 ($p > .25$).

Normative Considerations

The results indicate that subjects in both experimental conditions made unfavorable inferences about the welfare population. Though the two groups of experimental subjects made similar inferences, the degree of logical justification for the inferences differs sharply between the two groups.

Subjects in the typical sample condition would seem to be quite justified in their negative inferences. They read about a welfare recipient, described as typical at least with respect to her degree of dependence on welfare, whose life was a tangle of social pathology and personal irresponsibility. It is not unreasonable for subjects to infer from this information base that welfare recipients may be more incorrigible than they had thought.

In contrast, subjects in the atypical sample condition would seem to have had little justification for their unfavorable generalizations. First, they were given information about the average length of stay on welfare that was much more favorable than their prior beliefs (uninformed control subjects guessed that average length of stay on welfare was 10 years

and that the length of time required to remove 90% of recipients from the rolls was 21 years). The particular welfare case presented to subjects in the atypical sample condition was disturbing, to be sure, but subjects knew that the central figure was quite atypical in at least one important respect. The total package of information available to subjects in this condition amounted to the knowledge that there existed at least one irresponsible welfare recipient who had been on welfare about as long as they previously had thought was typical, plus the knowledge that the average recipient was on welfare for a much shorter time period than they had thought. This information would appear, on logical grounds, to be more consistent with favorable inferences about welfare recipients than with unfavorable inferences.

Failure of Statistical Information Alone to Influence Attitudes

The failure of the statistical information by itself to have any effect on attitudes should be noted. It will be recalled that the informed control subjects were provided with the favorable welfare statistics before responding to the dependent measures. Logically, it might have been expected that this information would have had a favorable effect on attitudes toward welfare recipients. Indeed, to refuse to change opinions in a favorable direction after finding out that lengthy stays on welfare are very much rarer than one had thought is tantamount to asserting the curious belief that average length of stay on welfare is irrelevant to an evaluation of the character and motives of recipients. Yet attitudes of the informed control group did not differ from those of the uninformed group, $t(45) = 1.02, p > .25$. This finding is similar to observations made by Nisbett and Borgida (1975; Borgida & Nisbett, 1977; Nisbett, Borgida, Crandall, & Reed, 1976) that statistical summary information often does not have the impact on inferences that normative considerations require. (It should be acknowledged, however, that the N for this comparison is not large and the risk of a Type II error is therefore high.)

It is important to note that it is possible that subjects *did* hold the belief that average length of stay on welfare is irrelevant to an evaluation of the character and motives of welfare recipients. If so, then control subjects were under no obligation to change their beliefs about recipients' character and motives when given information about the length of stay on welfare, and more importantly, experimental subjects would have been under no obligation to respond to sampling information. If the sampling information concerned a dimension they regarded as literally irrelevant, they would not be obliged to refrain from generalizing from the atypical sample. In Study 2 this interpretive problem was avoided by presenting subjects with sampling information regarding a parameter that was identical to the one subjects were asked to estimate.

Study 2

Study 2 was an attempt to replicate and extend the findings of Study 1 while avoiding two of the interpretive problems of that study. Study 2 provided a stronger test of the hypothesis that people are willing to generalize from biased samples than did Study 1. The sampling information in Study 2 concerned a dimension (length of time on welfare) that was logically highly related to the judgments that composed the dependent variables, but it was not identical to the dependent variables. In Study 2, the sample presented to subjects was described as either biased or unbiased with respect to the very population parameters that subjects were later asked to estimate. A second interpretive difficulty for Study 1 stems from the fact that the welfare case sample presented to subjects in Study 1 was highly consistent with a cultural stereotype about welfare recipients. It is possible that the counternormative inferences of the atypical sample subjects would be limited to such instances of stereotype confirmation. In Study 2 the nature of the sample was manipulated. For some subjects the sample was consistent with cultural stereotypes about the population in question, and for other subjects the sample sharply contradicted the cultural stereotypes.

Method

Overview

In Study 2 subjects rated their attitudes toward the population of prison guards in the U.S. A 2×3 experimental design with an additional control group was employed. All experimental subjects viewed a videotaped interview with an alleged prison guard. Half the experimental subjects saw a very decent and humane man posing as a guard; half the experimental subjects saw an inhumane, almost brutal, guard. Independent of the humanness manipulation, sampling information was manipulated within each of the humanness conditions. One third of the subjects were told that the guard they would see was highly typical and representative of those who worked in his prison, one third were told that the guard was quite atypical of those who worked in the prison, and one third were told nothing about the degree of typicality or representativeness of the guard. Control subjects saw no videotaped interview. Both experimental and control subjects completed a dependent measure questionnaire concerning attitudes toward prison guards. It was anticipated that subjects who saw the humane guard would express relatively favorable opinions about prison guards generally, that subjects who saw the inhumane guard would express relatively unfavorable opinions about prison guards generally, and that this would be true whether the guard was described as typical, as atypical, or whether nothing was said about the guard's typicality.

Materials and Procedure

Subjects (147 University of Michigan introductory psychology students of both sexes) participated in groups of 5-14. In all experimental conditions, subjects were seated in front of a 19-in. (.48-m) television monitor and were asked to read a cover sheet titled "Survey of Audio-Visual Materials." The cover sheet alleged that "the Higher Education Center of the Institute for Social Research is engaged in pretesting a new type of teaching aid for social science courses."

Videotaped interviews are being used more and more in social science research. . . . We have collected a number of videotaped interviews from a number of research projects that are being conducted at the Institute. We will show you several of these and ask you to evaluate them as to their interestingness and potential appropriateness as teaching aids. . . . The first interview you will see is an interview with a prison guard at a state prison in Michigan conducted by Dr. Nisbett, who has been studying different aspects of the prison system including prisoner-guard relationships, living conditions, and various aspects of a prisoner's life. During this time he had the opportunity to get to know many of the guards and

prisoners. He was able to interview nearly all of the guards at the prison, a total of about 60. The interviews took place in a very informal setting away from the prison. The researcher was well acquainted with these men by the time of the interviews and they felt quite comfortable with him. The guards also knew that the information would in no way get back to their superiors. It therefore seems clear that the views expressed by the guards are their own and that they felt free to be truthful.

The cover sheet ended with the sampling information described below. At each session, subjects were randomly assigned to one of the three sampling information conditions or to the control condition. Thus each group of subjects who viewed the videotaped sample contained some who believed they were seeing a typical prison guard, some who believed they were seeing an atypical guard, and some who had no information about the typicality of the guard they were seeing.

After watching the interview, subjects filled out a questionnaire allegedly assessing the interestingness and educational value of the interview. Then, under the pretext that ratings of interestingness might be affected by their attitudes about the penal system, subjects filled out a questionnaire assessing attitudes toward the system of justice in the U.S., penal institutions, and the critical dependent variable of attitudes toward prison guards. Finally, subjects responded to a manipulation check measuring their recall of the sampling information and were debriefed.

Manipulation of Humaneness of Sample

Half of the 108 experimental subjects saw an 8-minute interview with a young black male who presented himself as a particularly humane and decent sort of person who regarded his prison charges as fellow human beings worthy of his respect and concern. The other half of the subjects saw an 8-minute interview with a young white male who presented himself as a bitter, contemptuous man who regarded prisoners as subhumans: dangerous refuse to be controlled by harsh, possibly even brutal, means. A few excerpts will convey the flavor of the interviews.

The interviewer asked both men "what role the penal institution should play in society as far as rehabilitation goes." The humane guard responded by saying he believed that rehabilitation was the most important thing the prison system can do and that he believed it was institution's job to "help these guys out and get them back on their feet." The inhumane guard responded explosively: "Rehabilitation is a joke. . . . These guys are losers. The prison's job is to keep them away from society." In response to the interviewer's question about "whether these people belong here," the humane guard responded that "just like anywhere, there are some basically good guys here and some pretty

rotten ones. All of them broke the law, that's why they're here. But a lot of these guys are just ordinary people in bad situations . . . A lot of them didn't have jobs. They were broke. They got a few bad breaks and didn't know any other way out." The inhumane guard again exploded: "Of course they deserve to be here! These guys can't be bothered with going out and working. They want something and they take it. You let them out of here and they'll just go right back to what they've always done." The interviewer asked, "How would you define your job?" The humane guard responded: "Well, I have to make sure people keep in line, follow the rules. But I'm not here to give them a hard time. Part of my job is to help the prisoner put in his time and stay out of trouble." The inhumane guard said that his job was to "keep those guys in line, which isn't easy. I'm here to teach them a little respect for authority."

Sampling Information Manipulation

Typical sample conditions. Subjects in these conditions read, just before viewing the guard, that "The interview you will see is with a guard that Dr. Nisbett felt was one of the most typical guards in the prison. His answers seemed to be very representative of the ones given by the other guards who were interviewed."

Atypical sample conditions. Subjects in these conditions read that "The interview you will see is with a guard that Dr. Nisbett felt was one of the best and most humane (humane guard condition)/worst and least humane (inhumane guard condition) guards at the prison. His views, and his behavior at the prison, made him one of the three or four most/least qualified for the role, in Dr. Nisbett's opinion."

No sampling information conditions. Subjects in these conditions read only that "You are now going to see an interview with one of the guards."

Control, No Sample Condition

Thirty-nine control subjects saw no videotaped interview. These subjects were told that their attitudes about a number of political and social issues would be assessed. They then responded to the same dependent measures questionnaire presented to experimental subjects.

Dependent Measures

Subjects answered a number of questions about their attitudes toward the justice and penal systems, and at the end of this questionnaire were asked the four questions that constituted the chief dependent measures.

1. How important an aim do you think rehabilitation is for most guards? (1 = not at all important, 5 = extremely important.)

2. How fairly do you think guards treat the inmates? (1 = extremely unfairly, 6 = extremely fairly.)

3. How concerned do you think the average guard is with the welfare of the prisoners? (1 = not at all concerned, 5 = extremely concerned.)

4. How competent in his work is the average guard? (1 = extremely incompetent, 6 = extremely competent.)

A composite score for attitudes toward prison guards was formed by summing all four questions.

Manipulation Check

To determine whether subjects correctly recalled the sampling information, they were asked, after all other measures were completed: "How was the guard that you saw selected from among all those interviewed?" Answer alternatives were *most representative, one of the best, one of the worst, and don't know.*

Results

Manipulation Check

In the typical conditions, 86% of the subjects correctly recalled that the guard they saw had been selected because he was the most representative of those interviewed. In the atypical conditions, 89% of the subjects correctly recalled that the guard they saw had been selected as "one of the best" (humane guard condition) or "one of the worst" (inhumane guard condition). In the no sampling information condition, 94% of the subjects correctly indicated that they didn't know the basis on which the guard was selected.

Effects of Guard Humaneness and Sampling Information

Results for experimental conditions were analyzed by a two (levels of guard humaneness) by three (levels of sampling information) analysis of variance (ANOVA) of the summed guard items. Responsiveness to the character of the guard would be revealed by the main effect for the guard humaneness variable. Responsiveness to the sampling information would be revealed by the interaction between guard humaneness and sampling information: If subjects were responsive to the sampling information, then opinions about prison guards for subjects exposed to the humane guard should have been increas-

Table 2
Mean Attitudes Toward Guards as a Function of Guard Humaneness and Sampling Information

Group	Sampling information		
	Typical	No information	Atypical
Guard type			
Humane			
<i>M</i>	12.56	13.28	11.94
<i>SD</i>	2.09	1.90	2.62
Inhumane			
<i>M</i>	9.44	10.44	10.11
<i>SD</i>	1.85	2.43	2.03
Control			
<i>M</i>		10.97	
<i>SD</i>		2.67	

Note. The higher the mean, the more favorable the attitude toward guards.

ingly favorable as sampling assurances improved, and opinions for subjects exposed to the inhumane guard should have been increasingly unfavorable as sampling assurances improved.

As may be seen in Table 2, subjects were markedly responsive to the humaneness of the guard they saw and not responsive to the sampling information they were given. The humaneness of the guard accounted for 26.4% of the total variance in opinions about prison guards, $F(1, 102) = 38.43$. Information about sampling did not modulate this effect: The interaction accounted for only 1.2% of the variance ($F < 1$). Separate ANOVAs for each of the four items comprising the composite index in Table 2 revealed the same pattern—highly significant effects of guard humaneness for each question and trivial interaction effects for each question (all interaction $F_s \approx 1$).

The interaction term in the 2×3 ANOVA was not the most sensitive possible test of subjects' responsiveness to sampling information because it included the intermediate, no sampling information condition. Therefore, a simple 2×2 ANOVA of the summed scores was performed, retaining only the extreme typical and atypical sampling information

conditions. In addition, those few subjects who incorrectly recalled the sampling information were deleted from the analysis. The results of this more sensitive test do not alter the implications of the major analysis. The F for guard humaneness was still immense and accounted for 29.6% of the total variance. Planned comparisons showed that the effect of the guard humaneness was significant both when subjects were assured that their sample was typical (p for contrast between typical humane and typical inhumane conditions $< .001$) and when assured it was atypical (p for contrast between atypical humane and atypical inhumane conditions $< .05$). The interaction between guard humaneness and sampling information still accounted for only 2.3% of the variance, $F(1, 59) = 1.92, ns$.

Comparison With Control Group

A comparison of experimental subjects' views with those of control subjects who saw no videotaped interview is also instructive. Control subjects held a quite unfavorable view of prison guards as a group. The means for all four items were located in the decidedly unfavorable region of the scale. Thus the experimental subjects who viewed an inhumane guard saw a sample more or less consistent with their prior beliefs, whereas those who viewed a humane guard saw a sample contradicting their prior beliefs. Regardless of whether they saw a consistent or a contradictory sample, experimental subjects' beliefs about guards were altered. A one-way ANOVA of summed scores contrasting all humane guard conditions, all inhumane guard conditions, and the control condition was highly significant, $F(2, 144) = 17.07, p < .0001$. Neuman-Keuls contrasts showed that both the humane guard group and the inhumane guard group differed from the control group ($ps < .01$ and $.05$, respectively). It is thus clear that subjects were willing to generalize even from a sample that contradicted their prior beliefs.

Subject Assumptions About Population Variance

The failure of subjects to respond to obvious sample bias could be justified logically

under only one circumstance. Even a badly biased value is acceptable as a basis for parameter estimation if variance of the estimated dimension is known to be low. If subjects had believed that the population of prison guards was of a piece ("if you've seen one, you've seen them all"), they would have been justified in ignoring the bias in the atypical conditions. This belief would have been a dubious assumption at best, but in any case subjects did not have the belief. In a follow-up study designed to determine whether subjects believe that all guards are alike, 39 subjects from the same introductory psychology pool were asked to estimate, for a sample of 20 guards, how the best and worst guard would best be described in terms of the answer categories for each of the four dependent variable questions. (The best and worst of 20 guards corresponds roughly to the percentile values of best three or four and worst three or four of sixty guards). The presumed differences between best and worst guards were massive for each question. For example, follow-up subjects thought that the best guard would treat prisoners between "somewhat fairly" and "very fairly" (2.59 on a 6-point scale) and that the worst guard would treat prisoners "very unfairly" (5.03). The difference between best and worst guard was significant for this question and the other three at the .0001 level. Thus subjects did not presume that the guard population was extremely homogeneous and therefore were not justified in accepting an extreme value as a reasonable estimator of the population mean.

Discussion

Why were subjects so willing to generalize from samples of unknown typicality, and even to generalize from samples known to be atypical? It seems to us that the first consideration in answering these questions is to ask if subjects know, in the abstract, what generalizations are allowed under each of these conditions. If subjects are not even aware of any rules or guidelines to follow, it would not be surprising to find that they were willing to generalize. If, however, subjects are aware of the limitations imposed by the sampling

procedure, then we need to examine more closely the way this information is processed.

Consider first subjects' willingness to generalize from samples of unknown typicality. There are no rules stemming from formal statistical theory of inductive inference to govern generalization in such cases. Familiarity with formal theory provides only a rather vague injunction to be less confident of the implications of evidence drawn from samples of unknown typicality than of evidence drawn from samples known to be randomly drawn or known to be drawn from the center of the population distribution. Neither formal theory nor common sense implies that one should eschew *any* generalizations from samples of unknown typicality. We are exposed daily, in person and through the media, to samples of many populations—plumbers, Chicanos, gay militants, antiabortionists. To refuse to generalize at all from such haphazard samples would seem to be much too conservative, resulting in ignorance where some knowledge is possible. Clearly some caution is warranted, but precisely how much caution is required by comparison to the rare real-world case in which one can be sure that one's sample is typical? This is such a complicated question that formal inductivists have not even addressed it. It is therefore not particularly surprising to find the layperson generalizing as readily from samples whose representativeness is unknown as from samples known to be typical: There are no rules of thumb concerning precisely how much caution to exercise when confronted with uncertainty about sample representativeness. We hope that the research stemming from Kahneman and Tversky's seminal work on lay inductive practice will stimulate formal theorists to develop rules for complicated, real-world inductive problems, and that these can be communicated to laypeople in such a way as to guide the conduct of daily cognitive life (cf. Goldman, 1978; Nisbett & Ross, 1980).

Now consider subjects' willingness to generalize from samples known to be atypical. This seems to us to involve some additional, and different, considerations. No familiarity with formal theory of inductive inference is required to know, in the abstract, that one should not generalize from samples known to

be atypical. Nevertheless, subjects did so generalize. Lack of knowledge concerning what generalizations are allowable is therefore probably not the sole reason that subjects make unwarranted generalizations. We need to examine the cognitive processes that could produce generalizations even when subjects recognize their inappropriateness in the abstract.

If one were to assume that subjects knew what attitudes they had toward the relevant population of welfare recipients or prison guards before they ever arrived at the experiment, then it might be assumed that subjects could avoid any influence of the sample simply by calling up these attitudes and ignoring the information about the sample. There is no reason to assume, however, that subjects have anything like a precise record of previous attitudes toward the relevant populations, or much ability to recognize that their attitudes have been shifted by the experimenter's presentation of vivid evidence. Instead, work by Bem and McConnell (1970), Goethals and Reckman (1973), and Nisbett and Wilson (1977) suggests that people's attitudes often shift in response to new evidence and arguments without any awareness that a change has taken place. When subjects in the present experiments answer the dependent variable question about the populations, they might think they are disregarding the sample, but in fact it is likely that the sample already has influenced their attitudes quite unwittingly. One possibility is that such unwitting influence is memory-mediated. The vivid sample information probably serves to call up information of a similar nature from memory (cf. Bower, 1972; Collins & Loftus, 1975; Nisbett & Ross, 1980). These memories then would be disproportionately available when judgments are later made about the population.

We are proposing that subjects may engage in an unconscious, memory-mediated generalization from sample to population that remains unaffected by any conscious processing of information about sample typicality. Knowledge of the inappropriateness of generalization may have no effect on tendency to generalize because subjects cannot observe the process that produces generalization and

therefore cannot prevent it or correct for it. Evidence that unconscious inferences may fail to be corrected by conscious recognition of the biased processes that produce them comes from work by Lichtenstein, Slovic, Fischhoff, Layman, and Combs (1978). These investigators showed that the vividness and imaginability of events unduly influence subjects' estimates of their frequency and likelihood. Even when subjects are warned against this source of bias in their estimates, they show the bias in full strength. Subjects probably cannot directly observe the biased process in operation and cannot determine its precise influence on their judgments, thus they are unable to correct for it.

It seems likely that unwitting generalizations from atypical samples would occur in proportion to the vividness of the sample. Work by Wells and Harvey (1977) showed that, when information about typicality of samples is presented but not made to compete with vivid target information, people will use the sample base rate to a degree in their predictions about targets (although not to a normatively appropriate degree; cf. Borgida, 1978). Insensitivity to sampling considerations thus is probably more likely to occur when, as in the present research, the information about the sample is highly vivid and therefore likely to trigger unconscious, memory-mediated inferential processes.

We should note that there is at least one ecologically frequent source of sample bias to which people may be quite sensitive. Just as recent research has shown that people are responsive to base rates when making predictions if the base rates have a causal interpretation (Ajzen, 1977; Tversky & Kahneman, 1978), it seems likely that people are sensitive to sample bias when it can be given a causal interpretation. Most shoppers are wary of cellophane-wrapped packages of fruit in unfamiliar supermarkets. The possibility that the lovely strawberries on top may be a biased sample of the contents of the package will be salient because the motive for presenting such a biased sample will be obvious. Thus people should have some protection against their would-be manipulators when a causal interpretation for sample bias (self-serving motives) is available.

On the other hand, the present results provide little basis for optimism about people's defenses against biased samples. Instead, the results suggest that, at least when sample information is vivid and a causal interpretation for bias is lacking, people may make quite unwarranted generalizations from highly atypical samples.

References

- Ajzen, I. Intuitive theories of events and the effects of base rate information on prediction. *Journal of Personality and Social Psychology*, 1977, 35, 303-314.
- Bem, D. J., & McConnell, H. K. Testing the self-perception explanation of dissonance phenomena: On the salience of premanipulation attitudes. *Journal of Personality and Social Psychology*, 1970, 14, 23-31.
- Borgida, E. Scientific deduction—Evidence is not necessarily informative: A reply to Wells and Harvey. *Journal of Personality and Social Psychology*, 1978, 36, 477-482.
- Borgida, E., & Nisbett, R. E. The differential impact of abstract vs. concrete information on decisions. *Journal of Applied Social Psychology*, 1977, 7, 258-271.
- Bower, G. H. Mental imagery and associative learning. In L. Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, 1972.
- Collins, A. M., & Loftus, E. F. A spreading-activation theory of semantic processing. *Psychological Review*, 1975, 82, 407-428.
- Goethals, G. R., & Reckman, R. F. The perception of consistency in attitudes. *Journal of Experimental Social Psychology*, 1973, 9, 491-501.
- Goldman, A. I. Epistemics: The regulative theory of cognition. *The Journal of Philosophy*, 1978, 75, 509-524.
- Kahneman, D., & Tversky, A. Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 1972, 3, 430-454.
- Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, 4, 551-578.
- Nisbett, R. E., & Borgida, E. Attribution and the psychology of prediction. *Journal of Personality and Social Psychology*, 1975, 32, 932-943.
- Nisbett, R. E., Borgida, E., Crandall, R., & Reed, H. Popular induction: Information is not necessarily informative. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior*. Hillsdale, N.J.: Erlbaum, 1976.
- Nisbett, R. E., & Ross, L. D. *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, N.J.: Prentice-Hall, 1980.
- Nisbett, R. E., & Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 1977, 94, 231-259.

- Rein, M., & Rainwater, L. How large is the welfare class? *Challenge*, 1977, September-October, 20-33.
- Ross, L. D., Amabile, T. M., & Steinmetz, J. L. Social roles, social control, and biases in social-perception processes. *Journal of Personality and Social Psychology*, 1977, 35, 485-494.
- Sheehan, S. Profiles: A welfare mother in Brooklyn. *New Yorker*, 1975, 51, 42-46.
- Tversky, A., & Kahneman, D. Belief in the law of small numbers. *Psychological Bulletin*, 1971, 76, 105-110.
- Tversky, A., & Kahneman, D. Causal schemata in judgments under uncertainty. In M. Fischbein (Ed.), *Progress in social psychology*. Hillsdale, N.J.: Erlbaum, 1978.
- Wells, G. L., & Harvey, J. H. Do people use consensus information in making causal attributions? *Journal of Personality and Social Psychology*, 1977, 35, 279-293.

Received April 1979 ■