Journal of Experimental Psychology: General
1996, Vol. 125, No. 1, 52–68

0096-3445/96/$3.00

Inferring Category Characteristics From Sample Characteristics: Inductive Reasoning and Social Projection

Joachim Krueger and Russell W. Clement
Brown University

Inductive reasoning involves generalization from sample observations to categories. This research examined the conditions under which generalizations go beyond the boundaries of the sampled categories. In Experiment 1, participants sampled colored chips from urns. When categorization was not salient, participants revised their estimates of the probability of a particular color even in urns they had not sampled. As categorization became more salient, generalization became limited to the sampled urn. In Experiment 2 the salience of categorization in social induction was varied. When social categorization was not salient, participants projected their own responses to test items to members of a laboratory group even when they themselves did not belong to this group. When salience increased, projection decreased among nonmembers but not among members. In Experiment 3 these results were replicated in a field setting.

> The essence of knowledge is *generalization.* (Reichenbach, 1951, p. 5)

> Any empirical science begins with observation and generalization. (Hays, 1988, p. 4)

Opportunities to collect all the facts about a domain are rare. Biologists may know all surviving specimens of a vanishing species; senators may know all their peers. Many natural and social categories, however, can be understood only through the lens of limited experience. As Reichenbach (1951) contended, knowledge is possible in spite of incomplete information. The essential process through which it is acquired is generalization. Empirical knowledge is thus more than a catalog of facts, because it involves predictions of facts not yet experienced. To maximize confidence in such predictions, researchers typically define the target category carefully before drawing a sample of observations. They then face the question of "how does one . . . make general statements about the large body of potential observations, of which the data collected represent but a sample?" (Hays, 1988, p. 3).

Through generalization from instances, scientists and laypeople come to know their physical and social world. Ever since Aristotle (1846/1963) maintained that observation and inductive reasoning, rather than divine revelation, are the prime sources of human beliefs, empiricists have stressed the importance of learning through experience. British philosophers, among others, examined the properties and pitfalls of induction (Bacon, 1620/1960; Hume, 1748/1955; Mill, 1843/1974). Since then, empirical science has developed sophisticated statistical tools that guard against erroneous inferences, whereas intuitive induction largely proceeds with simple heuristics (Nisbett & Ross, 1980). Sometimes, intuitive induction matches statistical prediction (Gigerenzer & Murray, 1987), and sometimes it does not (Kahneman, Slovic, & Tversky, 1982). These differences have been thoroughly explored in tasks where target categories comprise the sampled observations. Little is known, however, about induction from sample data to other, nonsampled categories. Campbell (1957) cautioned researchers against sweeping generalizations: "There is always the possibility that the obtained effects are specific to the experimental population and do not hold true for the populations to which one wants to generalize" (p. 307). To learn more about cross-category induction, this research examines the effect of categorization itself on the spread of intuitive generalizations. The main thesis is that induction often spreads across category boundaries but becomes increasingly limited to the sampled category as the boundaries of the sampled category become more salient. The study begins with a review of relevant theoretical features of categorization and normative inductive inference. Three experiments are presented in which the salience of categorization varied. We conclude by discussing the relevance of the findings for scientific inference and intuitive social judgment.

Induction and Categorization

Normative models of induction stress that "the goal of sampling is to collect data from a representative sample

Joachim Krueger and Russell W. Clement, Department of Psychology, Brown University.

This research was supported, in part, by National Research Service Award 1 F31 MH10800-01.

We are indebted to Russ Church, Robyn Dawes, Richard Eiser, and Jack Wright for thoughtful comments on this work. Barbara Tannenbaum contributed to the development of the stimuli for Experiment 3, and Nancy Dunbar, Jacob Ham, and Erin Hurley assisted with data collection. Experiment 2 was presented at the Sixth Annual Convention of the American Psychological Society, Washington, DC, July 1994.

Correspondence concerning this article should be addressed to Joachim Krueger, Department of Psychology, Brown University, Box 1853, Providence, Rhode Island 02912. Electronic mail may be sent via Internet to joachim_krueger@brown.edu.

drawn from a larger population to make inferences about *that* [italics added] population" (Keppel, Saufley, & Tokunaga, 1992, p. 17). Induction is more complex, however, if the sampled category is related to other categories. In particular, the hierarchical organization of categories, the degree of overlap in features and exemplars, and the salience of categorization itself affect the spread of generalization.

Hierarchical Organization and Category Overlap

Categories can be organized into taxonomies with vertical and horizontal axes (Rosch & Lloyd, 1978). Porpoises and politicians belong to different species, but both are "living things." What are the consequences of hierarchical categorization for induction? Consider a sample of blue, flightless birds. Can observations derived from this sample be informative only about that species and no other? Certainly not. The feature *flightless* may apply to some other avian species and most other vertebrate classes, whereas the feature *blue* may apply to only a few other species. The hierarchical organization of categories means that some sample information from a specific category may generalize vertically and horizontally. Because vertical generalizations involve increasingly inclusive target categories, the informativeness of samples drawn from a subcategory tends to decrease. Horizontal generalizations can be made through inductive inferences about the superordinate category followed by deductive inferences about another subcategory. Rips (1975) showed that people engage this two-step process when the sampled category is typical of the superordinate category. Participants first inductively inferred the characteristics of birds from the characteristics of sparrows and then deductively inferred the characteristics of other avian species. Alternatively, when the sampled category is atypical, inferences can be analogical. The features of a neighboring category can be directly inferred from the available sample in the target category. In Rips's (1975) study, participants generalized characteristics of geese to similar species (e.g., ducks).

Not only features but also entire exemplars can appear in more than one category (Cantor & Mischel, 1979; Lingle, Altom, & Medin, 1984; Oakes & Turner, 1990; Smith, 1991). Some porpoises are domesticated pets or performers; some politicians have also been actors or convicted felons. The more categories overlap, the more likely it is that a sample contains exemplars that belong to multiple categories. With increasing overlap, inductive inferences may progressively reach beyond the boundaries of the sampled target category. What are the limits of such a progression? Often, there is no normative answer, because the precise degree of feature and exemplar overlap is unknown. Moreover, the sheer number of alternative categories is often unknown or even unknowable. In principle, any observation can be categorized in an infinite number of ways. There are, however, contextual factors that permit predictions of when and how far inductive generalizations spread across category boundaries.

Salience of Categorization

In generalizations to the sampled category, the sample parameters are the best estimators of the category parameters (Rorer, 1991). If exemplars belong to multiple categories, however, some of these categories may be more appropriate targets of induction than others. Most categories, though possible, are impractical or even bizarre (e.g., all things not brown; bearded Republicans from Rhode Island). Some categories are meaningful targets of induction because they are favored by social consensus, the situational context, or personal convenience. These factors may vary over time and across situations. They may affect the spread of generalization across category boundaries by making various categories more or less salient. Most important, the degree of salience of categorization itself may vary. When salience is low, category boundaries—by definition—are de-emphasized, and generalization may spread from the sampled category to the more inclusive categories and to neighboring categories of similar inclusiveness. When salience is high, category boundaries are sharply perceived. In the extreme, categories may be perceived as mutually exclusive when they are not. In this case, generalization may be limited to the sampled category.

In sum, the organization of categories in fuzzy hierarchies and context-dependent variations of category salience present predictable constraints on inductive inferences. It is a reasonable, admittedly inductive, hypothesis that these constraints are the same for different domains of knowledge, be they physical or social. Therefore, the present research examined the role of categorization in two distinct paradigms. First, a generic-induction paradigm uses chips and urns to convey information about samples and categories. This paradigm sacrifices the contextual richness of natural categories for the ability to derive precise predictions from a normative model of probabilistic induction. Second, a social projection paradigm uses self-related information as a vehicle for the formation of beliefs about social groups. Participants' own responses to test items serve as sample observations for generalizations about the responses of members of specific social groups or people in general.

Generic Inductive Reasoning

The purpose of a generic induction paradigm is to represent the critical features of the generalization process unambiguously (i.e., samples, categories, and assumptions about the relatedness of the categories). Statisticians often rely on the chips-and-urn paradigm to illustrate probability theory. Urns are metaphors for categories, and chips, for exemplars (Edwards, 1982). Drawn chips represent the sampled information, and estimates about the contents of urns represent inductive inferences. The appeal of the chips-and-urn paradigm is that it provides optimal values for statistical generalization. In a typical experiment, there are two urns. Suppose Urn U contains 80 blue and 20 red chips and that the alternative Urn A contains 20 blues and 80 reds. Also suppose that, a priori, each urn is equally likely to be sampled; that is, $p(\text{U}) = p(\text{A}) = .5$. The prior probability of drawing a blue chip is the prior probability of Urn U times the probability of drawing a blue chip, given urn U, plus the prior probability of Urn A times the probability of drawing a blue chip, given A: $p(\text{blue}) = p(\text{U}) \times p(\text{blue}|\text{U}) + p(\text{A}) \times p(\text{blue}|\text{A})$ (in this example, $.5 \times .8 + .5 \times .2 = .5$). Participants then draw a random sample of chips and estimate the probability that they have sampled from Urn U. If the sample consists of a single blue chip, Bayes's rule specifies the posterior probability of U as equal to the prior probability of U times the probability of drawing blue, given U, divided by the prior probability of blue. That is,

$$p(\text{U}|\text{blue}) = \frac{p(\text{U}) \times p(\text{blue}|\text{U})}{p(\text{blue})}$$

The probability that the blue chip was drawn from Urn U is $.5 \times .8/.5 = .8$. Bayes's rule stipulates that the probability of each of the two hypotheses (i.e., urns) changes after sampling, and so does the overall probability of the target characteristic (here: blue chips) in the experiment. Urn U, containing many blue chips, becomes more likely after sampling a blue chip, while Urn A, containing few blue chips, becomes less likely, and so the overall probability of blue chips increases. In this example, the posterior probability of blue, given that a blue chip was drawn, is $p(\text{blue}|\text{blue}) = [p(\text{U}|\text{blue}) \times p(\text{blue}|\text{U})) + (p(\text{A}|\text{blue}) \times p(\text{blue}|\text{A})]$; in this example, $.8 \times .8 + .2 \times .2 = .68$.

The estimation of the posterior probability of the target characteristic is more difficult when there are multiple distributional hypotheses. When decision makers estimate the probability of a characteristic in a category with unknown contents, their task is equivalent to estimating the contents of a single urn from sample draws. Instead of two hypotheses, there are 101 hypotheses about the probability (to the nearest .01) of a target characteristic. The prior probability of each Hypothesis H changes after sampling. After the draw of a blue chip, the probabilities of the hypotheses that many chips are blue increases, and the probabilities of the hypotheses that few chips are blue decrease. Bayes's rule specifies the posterior probability for each of the hypotheses. To calculate the overall posterior probability of drawing a blue chip, the posterior probability of each hypothesis needs to be multiplied by the probability of drawing a blue, given this hypothesis is true. These products are then summed: $p(\text{blue}|\text{blue}) = \Sigma[p(\text{H}|\text{blue}) \times p(\text{blue}|\text{H})]$. To adapt the generic induction paradigm for experimental study, three assumptions are made: uniform prior probabilities, independent sampling, and arbitrary categorization. Each assumption is now presented with a discussion of its psychological significance.

Uniform Prior Probabilities

The assumption of uniform priors means that all possible hypotheses are equally likely a priori. If nothing is known about the contents of an urn, the prior probability of each hypothesis concerning the percentage of blue chips is the same, namely $p(\text{H}) = 1/101 = .0099$. A psychological advantage of this assumption is that it captures the state of ignorance (Einhorn & Hogarth, 1985). A decision maker is ignorant if the degree of uncertainty is itself uncertain.[1] In social induction, it may not be possible to ascertain whether the prior probabilities are uniform. In a generic induction paradigm, however, the assumption of uniform priors can be stated explicitly, and optimal predictions can be computed (Krueger & Clement, 1994b, Experiment 3). This assumption also has a computational advantage: The formula for the revision of probability estimates is simply

$$p(\text{blue}|\text{blue}) = \frac{\text{number of blue chips drawn} + 1}{\text{total number of chips drawn} + 2} = \frac{k + 1}{n + 2},$$

where k is the number of successes (e.g., blue chips), and n is the size of the sample (Dawes, 1989; Krueger & Clement,

[1] To illustrate the difference between simple uncertainty and ignorance, suppose there are two urns. The contents of one urn are known to be 50 blue chips and 50 red chips: $p(\text{blue}) = .5$. The contents of the other urn are unknown, and the decision maker assumes that each possible probability of blue is equally likely. In this case, the probability of blue is .5 as well; that is, uniform priors: $p(\text{blue}) = (0 + 1)/(0 + 2) = .5$. The first case represents uncertainty with a single known prior probability, and the second case represents ignorance, where the expected value of the prior probability must be inferred from a set of uncertain hypotheses. Einhorn and Hogarth (1985) suggested that ignorance is psychologically less comfortable than simple uncertainty. Consistent preference of simple uncertainty over ignorance, they argued, would entail contradictory choice patterns. They proposed a thought experiment in which decision makers choose between Urn A, known to contain 50% blue and 50% red chips, and Urn B, where the proportions of blues and reds are unknown (i.e., uniform priors). Suppose participants expect a reward for drawing a blue chip, and they choose to sample from Urn A. This implies that the expected $p(\text{blue}|\text{B}) < p(\text{blue}|\text{A})$ and therefore $p(\text{red}|\text{B}) > p(\text{red}|\text{A})$. Because the probabilities in Urn A are specified, they may be less mutable than the unspecified probabilities in Urn B. This implies that $p(\text{blue}|\text{B}) < .5$ and $p(\text{red}|\text{B}) > .5$. If participants also choose to sample from Urn A when expecting a reward for drawing a red chip, the expected $p(\text{red}|\text{B}) < p(\text{red}|\text{A})$; that is, $p(\text{red}|\text{B}) < .5$. Therefore $p(\text{blue}|\text{B}) > p(\text{blue}|\text{A})$; that is, $p(\text{blue}|\text{B}) > .5$. Not to choose the urn with unknown odds, regardless of which color is rewarded, reveals superadditive and thus irrational choices; that is, $p(\text{red}|\text{B}) + p(\text{blue}|\text{B}) > 1$. We presented 78 participants with descriptions of Urns A and B. Sixty-two participants (79%) understood the assumption of uniform priors in Urn B. They estimated the probability of drawing a chip of a certain color to be .5. Participants were asked twice from which urn they would draw a chip. In one case, they would earn a (hypothetical) reward if a red chip were drawn; in the other case, they would earn a reward if a blue chip were drawn. Most participants chose Urn A (44 = 71%) in both cases. Fewer participants chose Urn B (10 = 16%) or declared indifference (8 = 13%) both times. Had they understood the equivalence of simple uncertainty and ignorance and had chosen accordingly, all should have consistently declared indifference, or, alternatively, chosen Urns A and B randomly. The observed choices violated the constraints of additivity.

1994b). If a sample consists of one blue chip, the posterior probability of blue is .67.

Independent Sampling

The assumption of independent sampling means that a sample is informative for the category from which it is drawn but uninformative for other categories whose contents have been assembled independently. A sample is also informative for a superordinate category consisting of the sampled category and other, independently assembled categories. Consider two independently assembled urns that contain equal numbers of chips. If a blue chip is drawn from one urn, the posterior probability of blue in the entire experiment (i.e., in the superordinate category) is the average of the posterior probabilities of blue in the two urns; that is, in the two subordinate categories: $p(\text{blue}|\text{blue}) = (.67 + .50)/2 = .58$.

Arbitrary Categorization

The assumption of arbitrary categorization holds when a category breaks up and its contents are randomly distributed across mutually exclusive and exhaustive subcategories. Then, samples obtained from the original category or from any of the resulting subcategories allow equally strong inductive inferences to all categories. Suppose the contents of a large urn are randomly divided into two small urns and that the assumption of uniform prior holds. It follows that the draw of a single blue chip results in an optimal posterior probability of .67 for each urn. It is irrelevant whether the chip is drawn from the original urn before its breakup (i.e., when categorization is not salient) or from one of several urns resulting from the breakup (i.e., when categorization is salient). A social psychological analogue of arbitrary categorization is the minimal group situation. When participants from the same pool are randomly assigned to two groups (e.g., the *blues* or the *greens*, Rabbie & Horwitz, 1969), the groups are similar and should be treated accordingly. However, participants' propensity to discriminate between arbitrarily created groups suggests that categorization, when made salient, overrides normative principles of induction.

Overview

This research examined the spread of inductive inference in a generic induction paradigm and in social categorization. In Experiment 1, participants learned about the three assumptions (uniform priors, independent sampling, and arbitrary categorization) and then sampled chips from urns. Optimal predictions follow from the assumptions and are compared with intuitive inductive inferences. The central hypothesis was that, in contrast to the Bayesian norm, a mere increase in the salience of arbitrary categorization would constrain the spread of induction across categories. In Experiment 2 we presented a test of this hypothesis in a minimal group situation. Participants were expected to view their own responses as sampling data and to project from themselves to their in-groups. The degree of projection to out-groups or to the superordinate population was expected to decrease when social categorization became salient. Experiment 3 was a test of this hypothesis in a field setting.

Experiment 1: Generic Induction

In Experiment 1 we adopted a chips-and-urn paradigm to examine induction for multiple hierarchically organized categories. The presentation of the design, the hypotheses, and the results are organized around the assumptions of uniform priors, independent sampling, and arbitrary categorization. At the outset, half of the participants were presented with two urns whose contents were assembled independently, and they sampled one of them. Within this group, salience varied with the continuity of the category boundaries over time. In Condition 1, the two urns merged into one after sampling. That is, the category boundaries were eliminated, and participants made estimates for the composite urn. In Condition 2, the urns did not merge, and participants made estimates about both. Because of the continued separation of the two urns, the salience of categorization was greater in Condition 2 than in Condition 1. The other half of the participants were presented at the outset with one large urn that later broke up arbitrarily into two small urns. Within this group, salience varied with the timing of sampling. In Condition 3, participants sampled from the large urn before it broke up into two small urns. That is, there were no category boundaries at the time of sampling. In Condition 4, participants sampled from one of two small urns after the large urn had broken up. Categorization was most salient in Condition 4 because participants witnessed the process through which the two urns were created.

Method

Participants and Procedures

One hundred ninety-seven Brown University undergraduates (58% women; mean age = 18.3 years) participated in exchange for credit in an introductory psychology course. They were told they would be participating in a study on human judgment. Participants were seated individually in private cubicles.

The experiment was performed in a session with other tasks that are irrelevant for the present purposes. The duration of a session was about 50 minutes. Macintosh IIci computers (Apple Computer, Inc., Cupertino, CA) controlled the presentation of the instructions and stimuli and the collection of data. Stimuli consisted of graphic depictions of urns, created in SuperPaint 3.0 (Aldus Corporation, San Diego). One large urn (8.0 cm × 5.6 cm) or two small urns (4.0 cm × 5.6 cm each) were presented. Below each urn there was a display of the number of chips it contained (i.e., 200 and 100 chips for the large and the small urns, respectively). To sample chips, participants clicked the mouse at a labeled button on the screen. With each draw, a circle with the letter *B* for blue or *R* for red appeared above the sampled urn.

Phase 1: Prior estimates. Participants read that "people make inductive inferences whenever they estimate the characteristics of a group of objects based on their knowledge of 'samples' of observations." Then, for half the participants, a graphic depiction of one urn appeared on the screen; for the other half, depictions of two urns, A and B, appeared. It was stated that large urns contained 200 chips and that small urns contained 100 chips. Chips could be either blue or red. The participants' task was to estimate the percentage of blue chips in each urn. Then the assumption of uniform priors was explained:

> Although you do not know the exact proportions of the colors in the urn(s), you know that in each urn, any combination of reds and blues is equally likely. There could be 100% reds or 100% blues. There could be 99% reds and 1% blues, or 99% blues and 1% reds, *or any combination in between.*

In the two-urn condition, instructions stated that the contents of the urns had been assembled independently. "Whatever the percentage of blue chips in Urn A may be, it tells you nothing about the percentage of blue chips in Urn B." After reading these instructions, participants estimated the percentage of blue chips in each of the presented urns.

Phase 2: Sampling and posterior estimates. A new screen presented a graphic depiction of one urn or two urns. In Condition 1, instructions stated that there was an "opportunity to randomly draw 10 chips from Urn A. To draw a chip, click the indicated button. A running count of chip color is provided in the upper left corner of the screen." With every click of the button, a circle with the letter *B* for blue or *R* for red appeared above the sampled urn. With each draw, a countdown of the number of remaining chips appeared underneath the urn.[2] Nine of the sampled chips were blue, and one was red. The serial position of the red chip varied randomly across participants. After drawing the tenth chip, participants learned that "all of the chips that you drew will now be returned to the urn." They clicked a button labeled *Return chips to urn* and witnessed one chip icon disappear at a time. Next, it was stated that "Now, all chips in Urn A and Urn B will be put into one large Urn C," and participants clicked a button labeled *Merge Urns.* The merging process appeared in a graphic display. The two urns moved toward each other until they blended into a single large urn. The label *C* replaced the two separate labels, and the contents were declared to be 200 chips. Then, participants were asked to "Please estimate again what percentage of chips are blue." They entered their posterior estimate and clicked a button labeled *Done.* Procedures in Condition 2 were similar, but the urns did not merge. Participants sampled chips from Urn A and made posterior estimates for both urns.

In Condition 3, a single large urn was presented in Phase 1, and participants sampled chips from it. In Phase 2, it was stated that "the chips from the urn are randomly distributed into two smaller Urns A and B, each of those now containing 100 chips." To start the breakup of the large urn, participants clicked a button labeled *Randomly Distribute Chips,* whereupon the image of the large urn gradually separated into two smaller urns, labeled *A* and *B.* Participants then made posterior estimates for both urns. Condition 4 was the same except that chips were sampled only from Urn A after the breakup of the large urn. Figure 1 displays the design of the study. The hands indicate the location and the timing of sampling. The optimal percentages are presented above each urn.

Results

Some participants (36%) had already taken part in Experiment 2. None of the median estimates in Experiment 1 differed depending on whether participants had also participated in Experiment 2 (all *ps* $> .4$).

Uniform Priors

Because normal distributions for the estimates seemed unlikely a priori, medians indexed central tendencies, and interquartile ranges (IQR, Sokal & Rohlf, 1969) indexed dispersion (the lower and upper bounds of the IQR bracket 50% of the observations). We evaluated differences in central tendencies with Wilcoxon's signed ranks test (Siegel & Castellan, 1988). Because there were multiple comparisons, the probability for acceptable error was set at $p < .01$. The first hypothesis was that participants understand the assumption of uniform priors. In Phase 1, in which no sample information was available, this assumption implied that the optimal prior percentage for the color blue was 50% in every condition (i.e., $(0 + 1)/(0 + 2) = .5$). Estimates were not expected to deviate from the optimal percentage. As predicted, all medians were identical to the optimum of 50%. The small dispersion (IQR = 5%) indicated that the distributions were unimodal.

Independent Sampling

The second hypothesis was that participants understand the assumption of independent sampling. The hypothesis was tested in analyses of the posterior estimates in Conditions 1 and 2. In Condition 1, after the sampling of Urn A and after the merging of Urns A and B, the optimal posterior percentage for A (from which chips were sampled) was 83.3% (i.e., $(9 + 1)/(10 + 2) = .833$); for B it was 50% because no chips were sampled. The optimal percentage of blues in the composite Urn C was the average of the two posterior percentages for A and B ($(83.3\% + 50\%)/2 = 66.7\%$). Consistent with the hypothesis, posterior estimates for Urn C (*Mdn* = 65%, IQR = 15%) were larger than the prior estimates ($z = 4.73, p < .001$) and did not differ from the optimum percentage ($z = 1.21$).

In Condition 2, the two urns did not merge, and participants estimated the posterior percentage of blue in the two urns separately. If participants understood the assumption of independent sampling, their percentage estimates should have been close to these optima. After sampling from Urn A, the optimum percentage of blues in this urn was 83.3%. Posterior estimates (*Mdn* = 80%, IQR = 25%) were larger than the prior estimates ($z = 5.79$, $p < .001$) and not different from the optimum ($z = 1.28$). The optimum posterior percentages for Urn B were identical to the priors (i.e., 50%) because this urn was not sampled and because the sample from Urn A was irrelevant for Urn B. The median posterior estimate was 50% (IQR = 20). These results suggest that independent sampling was understood.

[2] Note that the computation of optimal posterior estimates assumes sampling with replacement, whereas participants sampled chips without replacement. The reason for this discrepancy is twofold. First, unless populations are very small ($N < 20$), optimal posteriors vary little as a function of replacement. Second, the lack of replacement permitted participants to see the whole sample at one glance, thus reducing demands on working memory.

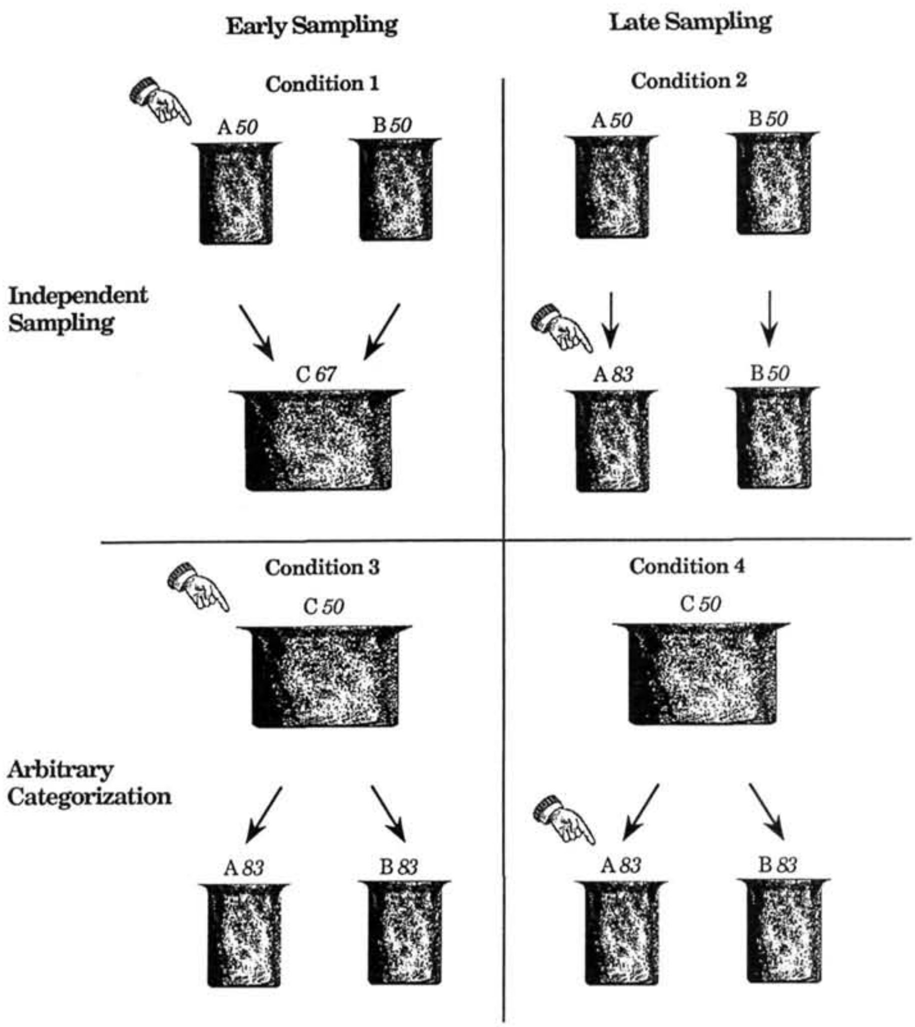

Figure 1. Images of urns, optimal percentage estimates, and timing of sampling.

Arbitrary Categorization

The third hypothesis was that the salience of categorization mediates the effect of arbitrary categorization. In Condition 3, the optimum percentage of blues was 83.33% for both the small urns that resulted from the random breakup of the large urn. The salience of categorization was low, because sampling occurred before the split. As expected, both median estimates were larger than the prior estimates (Mdn = 70%, IQR = 30% for A, and Mdn = 75%, IQR = 35% for B, both $ps < .001$). Both medians were lower than the optimum percentage (all $ps < .001$).

In Condition 4, participants sampled chips after the breakup, and thus the salience of categorization was high. The timing of sampling was irrelevant. Normatively, any sample from either Urn A or B was equally informative about both urns. If participants understood the arbitrariness of random categorization, all posterior estimates in these last two conditions should have been close to 83.3%. If, however, categorization was more salient when chips were sampled after the breakup of the original urn, participants may have revised their percentage estimates only for the sampled urn. As expected, estimates were larger than the priors only for Urn A, from which the sample was drawn (Mdn = 85%, IQR = 10%, $z = 5.44, p < .001$). The median did not differ from the optimum percentage ($z < 1$). For Urn B, from which no sample was drawn, many estimates (60%) were lower than the priors (Mdn = 25%, IQR = 60%). Although this difference did not reach the adopted significance level ($z = 2.30, p > .02$), it is noteworthy that of those participants whose estimates for Urn B were below 50%, most gave extremely low estimates (mode = 10%). Despite this surprising result, we can conclude that overall, the hypothesis was supported. Salient categorization constrained the spread of induction.

Discussion

Participants honored the assumption of uniform priors. In Phase 1, the median estimates did not differ from the optimum value of 50%. Participants also honored the assumption of independent sampling. In Condition 1 (Phase 2), in which the sampled urn merged with a nonsampled urn, posterior estimates were less extreme than in Condition 2, in which no merging occurred. Participants realized that increasing the size of the urn by adding chips of unknown color diluted the diagnosticity of the sample. Participants honored the assumption of arbitrary categorization when that categorization was not salient; that is, when the large urn was sampled before it broke up (Condition 3). If sampling occurred after the break-up (Condition 4), however, categorization was salient and, as predicted, estimates for blue increased for the sampled but not for the nonsampled urn. Indeed, most participants believed that blue chips had become less likely in Urn B. Possibly, the limits of salient categorization on the spread of inductive inferences are even stronger than hypothesized. Although the effect of negative generalization in Condition 4 did not reach the chosen level of reliability, it points to tantalizing possibilities. Perhaps participants did not understand that a sample from Urn A was informative about the probability of blue in the original urn from which both Urns A and B were derived. If participants did not revise their estimates for the overall probability of blue, increments in the estimates for Urn A after sampling would imply decrements in Urn B. Such reasoning is analogous to the "gambler's fallacy" (Tversky & Kahneman, 1971). Much as roulette players tend to become more confident that the color black will come up after a run of red, our participants may have expected that the probability of blue is diminished in one urn after it has been enhanced in the other.

Inductive Reasoning About People

The generic-induction experiment raises the question of whether its results replicate in social–cognitive contexts. How does the categorization of people into multiple social groups of varying inclusiveness affect the spread of generalization? Social induction is similar to generic induction because the same statistical rules apply. Individual observations of social behavior and personal characteristics are the sample data that allow inferences about group characteristics. Research on social stereotypes and projection has examined such inferences.

Stereotyping

To generalize the characteristics of individuals to the group to which they belong is one way to form a stereotype (Mackie, Allison, Worth, & Asuncion, 1992). If one wishes to learn what Italians are like, for example, a trip to Italy is instructive. After defining the group (e.g., Italians), sampling can begin (i.e., observing individual Italians). According to one model, stereotype learning and change are like bookkeeping (Rothbart, 1981). This model is Bayesian as it holds that stereotypes develop gradually and incrementally, following the influx of diagnostic sample information. Once a sample of observations is drawn, the question arises as to how far generalizations may extend. Is it appropriate to generalize to a parallel target category or to a superordinate, inclusive category? Can one generalize, for example, from a sample of Italians to the French or to Europeans? The results of Experiment 1 suggest that the salience of categorization limits the spread of generalization across social groups.

Rothbart (1981) suggested that inferences from stereotype-relevant information are constrained when that information is concentrated among category members that can be subtyped; that is, if the boundaries between a subordinate and a general category are salient. Consistent with this hypothesis, Weber and Crocker (1983) found that participants did not generalize from poor, Black corporate lawyers to corporate lawyers in general. Salient categorization may trigger perceptions of mutual exclusivity (Markman, 1989). Inasmuch as individual lawyers are seen as members of the category *poor and Black*, they may not be seen as members of the category *corporate lawyers*, although they logically belong to it (Rothbart & Taylor, 1992).

When categorization is not salient, induction flows more freely. Rothbart and Lewis (1988) presented participants with a description of a fraternity man. If he was typical of his group, a single behavior led to inductive inferences about the behavior among his fraternity brothers, among fraternity men on campus, and, to a lesser degree, among fraternity men in general. That is, when no distinctive features separate the target category from more superordinate categories, inductive inferences spread across the vertical axis of categorization.

Social Projection

When do inductive inferences spread across the horizontal axis of social categorization? In stereotyping, the main criterion of categorization within a level of inclusiveness is whether the self belongs to a category (in-group) or not (out-group; Simon, 1993). When making estimates about in-groups, information about the self is an omnipresent sample of one, and people generalize from themselves more to in-groups than to out-groups (Krueger & Zeiger, 1993; Mullen, Dovidio, Johnson, & Copper, 1992). There are three explanations for this asymmetry. According to the *bias explanation*, people project their own characteristics to their in-group and thus perceive a false consensus in that group (Ross, Greene, & House, 1977). By not projecting to their out-group, they contrast the out-group away from the in-group (Spears & Manstead, 1990).

According to the *Bayesian explanation*, it is reasonable to generalize to in-groups but not to out-groups if it is likely that the characteristics of the two groups are uncorrelated and if there is no information about other group members (Dawes, 1989; Hoch, 1987; Krueger & Zeiger, 1993). As Experiment 1 showed, sample information obtained from one group is not informative about the other when the assumptions of independence holds. The question is whether social categories are indeed independent. In the study by Mullen et al. (1992), students of varying ages and school affiliation constituted in-groups and out-groups. All groups were nested, however, within the category of *students*. In Krueger and Zeiger's (1993) research, the categories were *men* and *women*. Here, both groups were nested within the category of *people*. If members of two groups are similar to one another, there is feature overlap, and the characteristics of a single person (e.g., the self) may be informative for both groups. If, under these conditions, participants fail to generalize to an out-group that is similar to the in-group, they neglect the population base rates common to both. The neglect of base rates would be consistent with the bias of intergroup differentiation.

The present analysis suggests that, regardless of the actual similarities between groups, the spread of projection from self across groups may depend on the salience of social categorization. The more salient categorization is, the less likely should be projection to out-groups or to the general population. If this hypothesis is correct, it should be possible to control the degree of projection beyond the in-group by varying the salience of categorization while holding constant the degree of intergroup similarity. The minimal-group paradigm affords this possibility.

Minimal Groups

Tajfel (1970) noted that the ubiquitous ethnocentrism among social groups can be reproduced in the laboratory under minimal conditions of social categorization. Even if groups are created on a patently arbitrary basis, participants' thinking, feeling, and behaving favor groups to which they belong over groups to which they do not belong (for a review see Brewer & Kramer, 1985). Moreover, participants overestimate between-group differences on nonevaluative dimensions (Krueger, 1992; Krueger & Clement, 1994a). It seems reasonable to assume differences in the inductive reasoning of members and nonmembers of minimal groups. Members, more than nonmembers, may generalize (i.e., project) their own responses to others in the group. The minimal-group paradigm is attractive for the study of social induction for the same reasons that it is attractive for the study of in-group bias. Participants learn only to which group they belong but do not meet or expect any interaction with other members. Thus the paradigm controls—and limits—the information available about the groups and allows the researcher to assess the impact of a specific sample on inductive inferences. If there is no other information, participants can resort only to their own responses to a target stimulus. If they project their responses to the group or beyond, projection cannot be attributed to selective processing (exposure, attention, or memory) of information obtained from similar others (see Ross et al., 1977, for such explanations of the false consensus effect.)

The foregoing analysis of generic induction showed that the spread of generalization depends on the salience of categorization. Participants who sampled from a large category before it broke up into two small ones considered the sample informative for both small categories. In contrast, participants who drew samples after the breakup of the large category did not generalize beyond the small urn they had sampled. That is,a simple change in the order of sampling and categorization drastically reduced generalizations. In the minimal-group paradigm, the salience of social categorization can be manipulated by varying the order in which participants make percentage estimates for the general population and for a specific group (either an in-group or an out-group). When participants make estimates for a group after they make estimates for the population, the salience of categorization is low. In this condition, it may be evident that the population comprises members of different groups, and the degree of projection to the population and to the out-group may be high. In contrast, when participants make estimates for a group before they make estimates for the population, the salience of categorization is high, and projection may be limited to the in-group.

Experiment 2: Social Projection in Minimal Groups

The chips-and-urns scenario of generic induction was adapted, mutatis mutandis, to the area of social induction. Hierarchically organized social groups were the analogues of the urns. On the vertical axis, the general population was superimposed on specific groups. It constituted the analogue of the large composite urn. On the horizontal axis, arbitrarily created personality categories were created as social groups. They constituted the analogue of the small urns. Responses to various attitude statements were the analogue of the sampled chips, and making percentage estimates about the responses of others was the analogue of estimating the posterior probabilities of blue chips.

Participants made estimates in one of three orders. In Order 1, they made population estimates before being categorized into a group (low salience). In Order 2, they made population estimates after being categorized but before they made estimates about the group (intermediate salience). In Order 3, they made population estimates last (high salience). The first hypothesis was that overall, members would project more to the group than would nonmembers. The second hypothesis was that overall, participants would project more to a specific group than to the population. The third, and most crucial, hypothesis was that the degree of projection beyond the in-group (i.e., to the population or to another group) would decrease as the salience of social categorization increased. Categorization ostensibly reflected scores on a personality inventory but was in fact arbitrary. Next, participants indicated their agreement (or disagreement) with each of 10 statements and estimated the percentage of agreement with each statement. Finally, they rated the social desirability of each statement.

Method

Participants

One hundred forty-seven undergraduate students (55% female; mean age = 18.5 years) participated in exchange for credit in an introductory psychology course. The number of participants in each session ranged from 1 to 9.

Materials and Design

The sessions were held in the same laboratory, with the same equipment as Experiment 1. Assignment to groups was ostensibly based on performance on the "Wilson–Dobbs Personality Type Indicator." Participants actually took the Myers–Briggs Personality Type Indicator (Myers, 1962). The altered name ensured participants' naivité about the assessment. After being assigned to a personality group, participants read the following personality sketch of the typical member of one group:

> People in this category are enthusiastic, high-spirited, ingenious, imaginative and can do almost anything that interests them. They are quick with a solution for most difficulties, and are willing to help others with problems on their hands. They are particularly enthusiastic about books, they read or tell the parts they like the best to their friends. They are interested and responsive in class. They have warm, friendly personalities, but are not sociable just for the sake of sociability and seldom put their minds on possessions or physical surroundings. They are often quite organized, but sometimes rely on their spur-of-the-moment ability to improvise.

This sketch was ostensibly copied from the Wilson–Dobbs test manual. It was said to describe either the group to which the participant belonged or a different group. The sketch contained mostly positive, vague, and common statements taken from Myers's (1962) descriptions of various personality types. Most people consider statements of this type highly accurate descriptors of their personality (Forer, 1949; Sundberg, 1955). Participants estimated the percentage of people who may be described by the presented personality sketch, and they rated how accurately the sketch described them and how socially desirable it was.

The first independent variable was whether participants were assigned to the group described by the sketch. Because group members and nonmembers rated the same sketch, comparisons did not confound actual differences between participants and differences in group descriptions. The second independent variable was the salience of population and group estimates. In Order 1, participants made population estimates before being assigned to a group. In Order 2, they made population estimates after being assigned to a group but before making group estimates. In Order 3, participants made population estimates after they made group estimates. The within-subjects variable was the target category (group vs. population). Projection was measured through the responses to 10 statements drawn from the Minnesota Multiphasic Personality Inventory (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989; see Appendix A). Participants indicated their agreement (*yes* or *no*) with each statement, estimated the percentage of people in the group and in the population who would agree, and rated its social desirability.

Procedures

Order 3 will be used to describe the procedures. Participants first completed the "Wilson–Dobbs Personality Type Indicator." For each of 50 items, two response options appeared on the screen, and participants clicked the mouse at their preferred response. For example, when the question stem was "If you were a teacher, would you rather teach . . .," participants chose between *fact courses* or *courses involving theory*. After the response to the 50th item was recorded, the screen cleared and displayed the message "End of the Wilson–Dobbs Personality Type Indicator." Below this message, the words *Computing score . . .* and *Dimension K:* appeared. Next to these words, a series of numbers flashed on the screen, increased by 1 with each flash and stopped at 15. This process was repeated on subsequent lines with *Dimension H* (score = 23), *Dimension S* (score = 5), and *Dimension Q* (score = 16). The total score, 59, was displayed as the sum of the four individual scores. All participants received the same scores and were told that their sum score permitted their categorization into one of several personality types. By random assignment, participants were told they belonged to the *rational–intuitive* or to the *cognitive–empathetic* group.

After categorization, the personality sketch was presented. Ostensibly, the sketch described either the rational–intuitive or the cognitive–emphatic group. Each participant read the same sketch and, depending on the preceding categorization, the participant either believed that he or she was or was not a member of the described type. No nonmember expressed surprise about having to read a sketch that, ostensibly, did not describe them.

There were three category-level judgments: "How well does this category describe you? (1 = *not at all descriptive*, 9 = *very descriptive*); "What percentage of the population belongs to this category?" and "How socially desirable is it to belong to this category?" (1 = *very undesirable*, 9 = *very desirable*). The order of judgments was randomized across participants. Participants were then presented with 10 statements, drawn from the MMPI-2, and asked to "think of people who are rational–intuitive [cognitive–empathetic]. Of these, estimate what percentage would agree with the following statement." During a second presentation of the statements, participants were instructed to "think of the adults in the overall population. Of these, estimate what percentage would agree with each statement." Participants then indicated their own endorsements (agree vs. disagree) and rated each statement for social desirability. Finally, participants stated whether they had heard of the "Wilson-Dobbs Personality Type Indicator." If they had, they were asked to detail what they knew. All participants summarized what they felt was the purpose of the experiment.

Results

Four participants claimed they had heard about the "Wilson-Dobbs," but none realized that the test was actually the Myers–Briggs, nor did anyone present reasons to be excluded from analyses. Perceptions of the purpose of the experiment revolved around issues in personality assessment. Most participants found the task engaging, and many requested further information about the personality types. The group labels (*rational–intuitive, cognitive–empathetic*) had no effect on the dependent variables and are henceforth ignored. Separate 2 (status: member vs. nonmember) × 3 (salience: three orders) × 2 (sex) analyses of variance (ANOVAs) were performed on the category-level measures. The group assignment manipulation was successful. Group members considered the personality sketch to be a more accurate description of themselves (M = 7.26, SD = 1.42) than did nonmembers (M = 5.58, SD = 2.09), $F(1, 135) = 31.38$, $p < .001$.[3] Men estimated that a smaller

[3] This finding strikingly demonstrates the *Barnum effect*. In a typical Barnum experiment, participants rate how accurately a personality sketch describes them. Accuracy ratings are particularly high when the experimenters claim the sketch was written especially for the participant. Ratings are somewhat lower when the sketch purportedly describes most people (Snyder & Shenkel, 1976).

percentage of the population (M = 25%) belonged to the target group than did women (M = 32%), $F(1, 135) = 9.94$, $p < .01$. No other effects were statistically reliable for the category-level measures.

Social projection was indexed by point–biserial correlations between endorsements and consensus estimates. These within-subjects correlations were computed across items and separately for estimates about the target group and about the population. Correlations were then submitted to r-Z-r transformations (McNemar, 1962). Figure 2 shows the results. Z scores were analyzed in a 2 (status) × 3 (salience) × 2 (sex) × 2 (target: group vs. population) ANOVA with repeated measures on the last variable. There were no sex differences, and thus this variable was ignored in subsequent analyses. Differences between individual cell means were examined with Tukey's honestly significant difference test.

Consistent with the first hypothesis, group members projected more (mean r = .49, SD = .22) than did nonmembers (mean r = .36, SD = .35), $F(1, 140) = 10.90$, $p < .001$. Consistent with the second hypothesis, projection to the group (mean r = .47, SD = .45) was stronger than projection to the population (mean r = .38, SD = .49), $F(1, 140) = 6.87$, $p < .01$. More important, the two main effects were qualified by an interaction between status and target, $F(1, 140) = 17.90$, $p < .001$. The difference in projection between members and nonmembers occurred when the group but not the population was the target category. Supporting this analysis, the data in Figure 2 show that members projected reliably to the group in all order conditions (black bars), whereas nonmembers projected to the group only in Order 1. The white bars indicate that there were no differences between members and nonmembers in the degree of projection to the population.

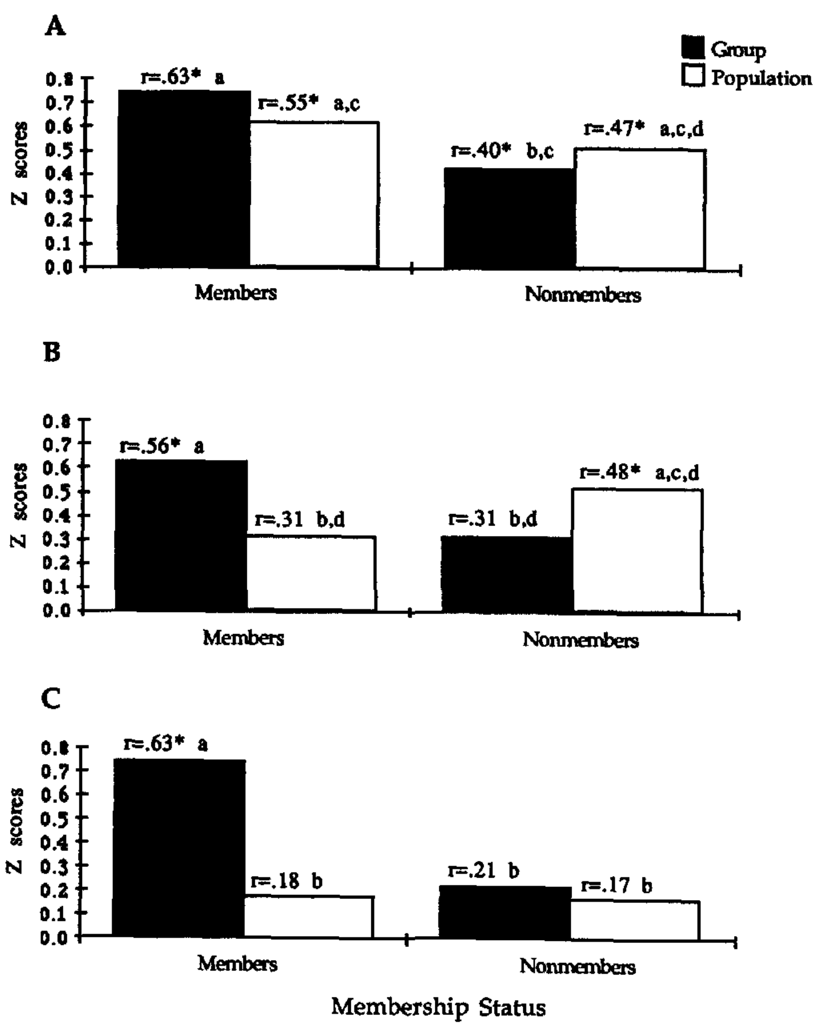

Figure 2. Mean within-subjects indices of social projection to the group and to the population. A: In Order 1, population estimates were made first. B: In Order 2, population estimates were made after group assignment. C: In Order 3, population estimates were made last. *Note.* *$p < .004$ (two-tailed). Different letters indicate differences at familywise error of $p < .05$ (Tukey's honestly significant difference).

Consistent with the third hypothesis, the reliable effect of salience indicated that projection was greater in Order 1 (mean $r = .51$, $SD = .30$) than in Order 3 (mean $r = .32$, $SD = .29$), $F(2, 140) = 6.64$, $p < .01$. Projection in Order 2 was intermediate (mean $r = .42$, $SD = .39$) and not reliably different from Orders 1 or 2. It appears that social categorization became increasingly salient across orders, and as a consequence, the spread of projection beyond the in-group decreased. As predicted, projection was reliable among group members and nonmembers when estimates about the population were made first (Order 1, see Figure 2). Projection was reliable only among group members when estimates about the population were made last (Order 3). The only other reliable effect was the interaction between salience and target, $F(2, 140) = 3.12$, $p < .05$.[4]

It is conceivable that the differences in social projection were mediated by the social desirability of the test items. On average, endorsements were related to desirability ratings (mean $r = .28$, $SD = .45$), and desirability ratings were related to percentage estimates (mean $r = .32$, $SD = .36$). To examine whether social desirability ratings played a role as a confounding third variable, the correlations for projection were computed again as partial correlations that controlled for the social desirability ratings. The Z scores for the zero-order and partial correlations were subjected to a joint 2 (status) $\times$ 3 (salience) $\times$ 2 (target) $\times$ 2 (correlation type: zero-order vs. partial) ANOVA with repeated measures on the last two variables. All the effects of the original ANOVA were replicated and were not qualified by interactions with correlation type. Overall, however, partial correlations (mean $r = .37$, $SD = .36$) were smaller than zero-order correlations (mean $r = .43$, $SD = .35$), $F(1, 140) = 17.00$, $p < .001$. Thus, social desirability did not moderate any effects between conditions, but it accounted for a small but reliable portion of social projection.

The foregoing analyses focus on variations in the mean level of projection depending on the salience of categorization. The salience hypothesis also suggests that the degree of projection to the group and the degree of projection to the population should be correlated across participants when salience of categorization is low but not when it is high. To test this idea, the Z scores for projection to the group were correlated with the Z scores for projection to the population. This was done within each of the six conditions. The participants for whom social categorizations was least salient were nonmembers who made estimates about the population first (Order 1). As expected, there was a reliable positive correlation: that is, $r(26) = .45$, $p < .05$. In contrast, social categorization was most salient for group members who made estimates about the population last (Order 3). For these, there was a negative correlation, $r(20) = -.40$, $p < .05$. In the other four conditions, correlations did not differ reliably from 0. These results show that the more the in-group was perceived to be an adequate target category for projection, the less the population was seen as an adequate target.

Discussion

Group members projected more to the group than did nonmembers, and both projected equally to the population. This finding underscores the analogy between generic induction and social projection. Participants generalized from sample information most strongly to the category from which the sample was drawn. They also projected to the out-group and to the population when they made population estimates first (Order 1). However, they projected little to the population or the out-group when they were categorized first (Order 3). This finding supported the hypothesis that projection (i.e., induction) beyond the in-group decreases when social categorization, however arbitrary it may be, becomes salient. Unlike in Experiment 1 (Condition 4), the salience of categorization in Experiment 2 did not lead to negative generalizations to the category that had not been sampled (i.e., the out-group). It is possible that such an effect would occur with further increases in the salience of social categorization. Variations in the order of the ratings of arbitrarily created laboratory groups may not have been a sufficiently strong manipulation for this purpose. As previously stated, however, in-group–out-group differences in projection to real social groups typically do not involve negative projection to out-groups (e.g., Krueger & Zeiger, 1993). In the present experiment, social categorization was arbitrary from the experimenters' perspective but not necessarily from the participants' perspective. After all, participants believed they were categorized depending on their score on a bona fide personality inventory. Most minimal-group studies on in-group favoritism follow similar procedures. It remains to be seen if the differences in social projection are obtained when groups are created in the most blatantly arbitrary fashion (e.g., by lottery).[5]

The differences in the salience of social categorization between the three order conditions in Experiment 2 were considered to be an experimental analogue of group formation. In the social world, people interact over time, and form groups gradually. As experience with an in-group increases, membership should become salient, and a sense of membership in superordinate, inclusive categories should be-

[4] Measures of the truly false consensus effect (TFCE) were computed by correlating each participant's endorsements with the differences between estimates and actual base rates across items (Krueger & Zeiger, 1993). The Z scores for simple projection and TFCE were subjected to a joint 2 (membership status) $\times$ 3 (order) $\times$ 2 (target) $\times$ 2 (projection measure: simple vs. TFCE) ANOVA with repeated measures on the last two variables. The effect of projection measure was reliable, indicating that the average TFCE score (mean $r = .34$, $SD = .39$) was smaller than the average simple projection score: mean $r = .43$, $SD = .35$), $F(1, 140) = 42.95$, $p < .001$. Aside from this effect, the pattern of TFCE mirrored that found for simple projection.

[5] In-group–out-group differences in social projection recently replicated in a within-participants design. Participants were arbitrarily categorized after taking the embedded-figure test or the rod-and-frame test, and each participant made percentage estimates for the in-group and the out-group (Clement & Krueger, 1995).

come less salient. Experiment 3 was designed to replicate the principal results of Experiment 2 in a longitudinal field study. Specifically, we predicted that as group membership (and thus social categorization) became more salient, members would maintain their level of projection to the group but would lessen their level of projection to the general population. No such effect was predicted for participants who did not belong to the group.

Experiment 3: Social Projection in a Longitudinal Field Study

The reduction of projection over time may depend on the confluence of two conditions. First, participants need continuing and deepening experience with group membership and thus with social categorization. Second, they need to make percentage estimates for the population rather than the in-group. If only one of these conditions exists, the degree of projection should be stable. Participants completed a 20-item questionnaire and estimated the responses of others. Participants in the experimental condition were enrolled in a popular course on persuasive communication. Students in this course get to know each other well and develop a common sense of purpose. They meet five times a week and sometimes more frequently. They deliver multiple speeches to the class and critique and commend each others' efforts. Typically, they consider this course a central experience of their college careers. Participants in this condition made percentage estimates for the population of college students. We expected that the degree of projection would be high at the beginning of the term (Time 1) and would diminish toward the end of the term (Time 2).

The test of the experimental hypothesis required two conditions. In one (group control), participants had the experience of group membership in the communication course but made estimates for that group rather than for the population. We expected the degree of projection to remain high at Time 2. In the other condition (population control), participants were members of a lab section of an introductory personality course in the psychology department. These students did not experience the type of group formation and team spirit that the communication students experienced. Attendance of lab sections was optional. Students met only once each week in a casual atmosphere. They spoke little and informally. Over the course of the term, some students changed labs and occasionally skipped meetings. They made estimates for the population. We expected the degree of projection to be equally high at Time 1 and Time 2.

Method

Participants and Stimuli

Ninety-two undergraduate students (59% female, average age = 21.1 years) volunteered to participate. Seventy-nine of these were recruited in a course on persuasive communication, and 13 were recruited in a lab section of an introductory personality course.

One of the instructors of the communication course and the first author wrote 20 statements. Five of these referred to perceived skill in public speaking (e.g., "I am familiar with the main principles of rhetoric"), five referred to emotions felt during public speaking (e.g., "I feel nervous before and during a public speech or presentation"), five referred to a speaker's sensitivity to the audience (e.g., "Before giving a speech, it is more important to clarify what one wants to say than what the audience is ready to hear"), and five were unrelated to public speaking (e.g, "I am neither gaining nor losing weight"). The instructors of the communication course and the present authors agreed that the statements referred to personal perceptions and evaluations and were also meaningful to students not enrolled in communication courses.

Design and Procedures

The design consisted of three conditions and two assessments 8 weeks apart. Most students in the communication course ($n = 56$) were in the experimental condition. They rated their agreement with each statement on a scale from 1 (*strongly disagree*) to 9 (*strongly agree*), and they estimated the percentage of college students who would give an agreement rating higher than 5. The group control condition comprised communication students ($n = 23$) who made percentage estimates for their own course instead of the population of college students. They responded to this request: "Please *estimate the percentage of students taking this class on persuasive communication that would give a rating higher than 5* on the rating scale." The population control condition comprised psychology students ($n = 13$) who completed the same questionnaire as the participants in the experimental condition. All participants closed by giving their age, sex, and the last four digits of their mother's phone number. This code number permitted to match participants' forms from Time 1 and Time 2 without sacrificing anonymity.

Results and Discussion

Ratings obtained in the communication course indicated that students made progress during the semester. The means changed reliably from Time 1 to Time 2 for 12 items, and the direction of each change was as expected, indicating greater perceived skill and less nervousness in public speaking (all $ps < .03$ or better; see Appendix B). Difference scores, M(Time 1) − M(Time 2), obtained in the experimental condition and in the group control condition were equivalent, as shown by the high correlation across items: $r(10) = .84$. No changes were expected or found among the psychology students (population control condition). The difference scores obtained in this condition were correlated neither with the difference scores in the experimental condition, $r(10) = .07$, nor with the difference scores in the group control condition, $r(10) = .13$.

We predicted that in the experimental condition, the level of projection to the population would decrease as a function of an increase in the salience of social categorization and independent of systematic changes in the group's characteristics. Appropriate items for a test of this prediction were items without changes in the mean rating from Time 1 to Time 2. The items that showed change among communication students reflected the success of the course and were inadequate for the analysis of projection. If, for example, communication students became less nervous during the term, their ratings of statements addressing nervousness might change. If they realized that other communication students, but not students in general, experienced a reduction in nervousness, their estimates about the population may have remained the same. Thus, if self-ratings changed but population estimates did not, correlations of projection would drop. Only the eight statements whose mean ratings were stable in all three conditions entered the test of projection. Limiting the analyses of projection to these items avoided the potential confound between actual change and the degree of projection. As shown in Appendix B, the standardized mean differences (Cohen, 1988) were smaller for the selected (all *d*s $< .10$) than the discarded items (all *d*s $> .10$).

We computed the within-subjects correlations between own responses and percentage estimates of agreement across the eight selected items with stable ratings. Figure 3 shows that the mean *Z* scores resembled those in Experiment 2. Projection to the population declined when the salience of social categorization increased (experimental condition: $t(56) = 2.55, p < .02$). There were no changes in projection when estimates were made for the in-group (group control condition) or when social categorization did not become more salient over time (population control condition).

Whereas the analysis of the selected items provided an unambiguous test of the main prediction, it is worth noting that the pattern of results also held for the discarded items. For these, projection in the experimental condition dropped sharply (mean *r*s = .53 [*SD* = .66] and −.06 [*SD* = .47] for Time 1 and Time 2, respectively). Recall, however, that this reduction confounds the increased salience of social categorization with actual changes in the group relative to the population. In the group control condition [mean *r*s = .64 (*SD* = .42) and .62 (*SD* = .56)] and in the population control condition [mean *r*s = .70 (*SD* = .29) and .71 (*SD* =.30)], the degree of projection was stable on these items.

Finally, the data in Experiment 3 permitted a test of whether there were stable individual differences in the degree of projection. Between the two assessment times, the individual *Z* scores of projection (based on the eight selected items) were correlated across participants [$r(54) = .46$, $p < .01$ (experimental condition), $r(11) = .25$, *ns* (population control condition), $r(21) = .39$, $p < .05$ (group control)]. Participants who projected more than others at the beginning of the term also projected more than others at the end.

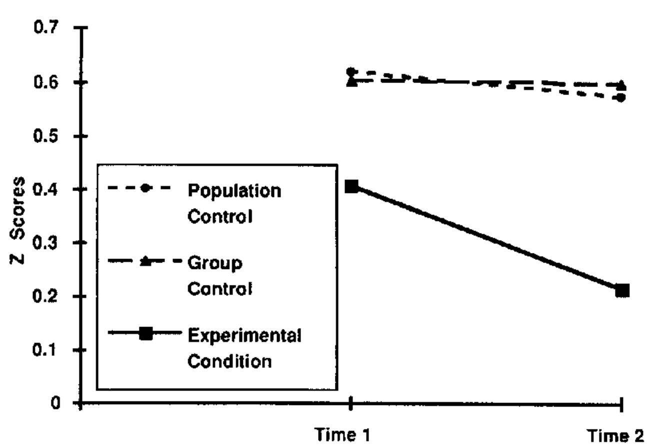

Figure 3. Social projection in a field setting.

The data in Experiment 3 not only duplicated the pattern found in Experiment 2, but in the relevant conditions, the mean correlations were nearly identical. The summary in Table 1 shows that in both experiments, projection to the in-group was equally high and unaffected by whether the distinction between levels of social categorization (group vs. population) was salient. When social categorization was salient, projection to the population diminished in both experiments. In neither experiment, however, was there the kind of negative induction observed in one condition of Experiment 1.

General Discussion

The simplest requirement of induction is that sample information should guide inferences about the characteristics of the sampled category. In all three experiments, estimates about the probability of a target characteristic were related to the probability of the characteristic in the sample. Participants revised the probability of blue chips in the sampled urn, and they projected their own responses to members of ad hoc and real-world in-groups. In Reichenbach's (1951) terms, participants recognized the predictive value of observations for that which had not yet been observed.

A more complex requirement of induction is that under certain conditions, sample information should guide inferences about the characteristics of categories larger than the sampled category (vertical generalization). In generic induction, participants honored the assumption of independent sampling. They revised the probability of blue chips in a large urn even when they had sampled only from one subset of that urn. In social induction, participants projected to the population when they had not yet had experience with a less inclusive in-group.

The least intuitive requirement of induction is that sometimes sample information should guide inferences about the characteristics of a category from which the sample was not obtained (horizontal generalization). Participants honored the assumption of arbitrary categorization when that categorization was not salient. When they had sampled from a large category, they increased probability estimates for blue chips for both small categories that emerged—through random splitting—from the large category. Similarly, in social induction, participants projected to an out-group when they made their estimates right after they had made estimates for the population (Experiment 2, Order 1). Participants may have remembered that the out-group belonged to the same population to which the in-group belonged.

The central prediction was that the spread of generalization would be constrained when categorization was salient.

Table 1
Mean Correlations of Projection in Experiments 2 and 3

	Vertical categorization			
	Nonsalient		Salient	
Target category	Exp. 2	Exp. 3	Exp. 2	Exp. 3
In-group	.63[a]	.60[c]	.63[b]	.60[d]
Population	.55[a]	.49[e]	.18[b]	.22[f]

Note. Exp. = experiment.
[a]Order 1. [b] Order 3. [c] Group Control Time 1. [d] Group Control Time 2. [e] Average of Experimental Condition Time 1 and Population Control Time 1 and 2. [f] Experimental Condition Time 2.

Consistent with this prediction, the generic-induction paradigm showed that participants abandoned the assumption of arbitrary categorization when they sampled a small category that had resulted from the random breakup of a large category. Probability estimates for blue chips increased only in the sampled but not in the nonsampled urn (for which, in fact, estimates decreased). Similar results emerged in the social projection paradigm. Here, participants projected little to an out-group or to the population after they had garnered experience with the in-group. In the eyes of the participant, the population may have turned into an out-group, consisting of all people minus the members of the in-group. Projecting to the in-group, but not projecting to the population or a specific out-group, creates a perceptual intergroup contrast that may enhance a person's social identity (Spears & Manstead, 1990).

In discussing the implications of these findings, it is useful to return to the introductory point that inductive inferences are critical ingredients of both scientific and intuitive reasoning. In formal scientific induction, the generalizability of data across category boundaries poses the question of external validity. In intuitive social induction, the generalizability of person impressions across group boundaries poses the question of intergroup stereotyping.

Estimation of External Validity

As statisticians and philosophers of science have noted (e.g., Hays, 1988; Reichenbach, 1951, see introductory quotes), knowledge accumulates through generalization from instances to populations. Researchers seek to ensure the reliability and internal validity of their inductive inferences by relying on normative statistical tools. The scientific method—or methods—guides inferences about the categories from which the samples were drawn.

It is neither practical nor desirable, however, to limit inductive inferences to the sampled categories. Studies on animal behavior underlie and improve studies of human psychology. Studies on college sophomores are cited to support conclusions about people in general. Studies conducted in the United States or in Europe find readers around the world. By exploiting available and convenient samples (and thus the populations to which the samples belong), science is opportunistic, gambling on the external validity of the findings. Campbell (1957) suggested that researchers should routinely ask "To what populations, settings, and variables can [the obtained] effect be generalized?" (p. 297). Unfortunately, the scientific method offers little help in specifying the optimal degree of generalization across the boundaries of the sampled category. In Experiment 1, the specification of optimal generalization across category boundaries came at the cost of presenting very explicit and detailed assumptions about categorization and sampling. Often, such assumptions cannot be made, and the optimal degree of generalization remains unknown. Campbell (1957) discussed several quasi-experimental corrections for possible selection bias. Somewhat pessimistically, however, he concluded that external validity can be ensured only by "defining the universe of reference in advance and selecting the experimental and control groups from this at random. [This] would guarantee representativeness if it were ever achieved in practice" (pp. 307–308). More recently, Abelson (1995) made a similar recommendation. He noted that it is difficult to generalize social psychological effects across different contexts when the experimental designs specify only a few contexts as fixed effects. As a remedy, he suggested to treat contexts as random factors with multiple levels. Both these recommendations constitute a return to the conservative idea that confident inductive generalization may target only the sampled categories. Sampling across the universe of categories obviates the need for generalizations across boundaries.

The question of generalizability of data to other populations thus remains murky. Because normative solutions are often lacking, researchers tend to either ignore or polemicize the problem. Some "human" psychologists deride animal research. Some life-span psychologists deplore the reliance of experimentalists on sophomore participants. Whereas the present research does not offer normative rules for generalization across categories, it highlights one mechanism that affects people's perceptions of how much generalization is justified. As the salience of categorization decreases, the inferential leap increases. Thus, inferences drawn from a given set of research data may appear thoroughly convincing or highly suspect, depending on whether attention is focused on the boundaries of the category from which the data were sampled.

Reduction of Social Stereotyping

The same principle applies to social stereotyping. The less salient social categorization is, the greater is the generalization from instances observed in one group to other, nonsampled groups. Inasmuch as perceptions of intergroup differences are a key feature of stereotyping, the role of salient social categorization is paramount. The minimal group paradigm was originally devised to examine the evaluative intergroup differentiation. Categorization is usually salient in this paradigm as participants are assigned to groups at the outset of the experimental procedures. Then they rate members of their own or the other group, and, typically, they attribute more favorable characteristics to the in-group than to the out-group (Howard & Rothbart, 1980). Similarly, under conditions of high salience of categorization, participants in the present research strongly differentiated between in-group members and others (out-groups or the general population). This differentiation emerged as a discriminative pattern of induction (or projection) above and beyond the effects of evaluative in-group bias. When the salience of social categorization was low, however, participants perceived the characteristics of an out-group and the characteristics of the general population to be similar to their own.

These results suggest that in-group bias may be reducible through decreases in the salience of social categorization. When participants rate the general population before they rate specific groups, out-group stereotypes may, at least momentarily, improve. Under this condition, participants may recognize the arbitrariness of categorization and rate out-group members as favorable as in-group members. Gaertner, Mann, Murrell, and Dovidio (1989) showed that the attenuation of in-group bias is unlikely to be accomplished through a reduction of perceived in-group favorability but rather through an improvement in perceived out-group favorability. A finding such as this is consistent with our finding that projection by out-group members, but not by in-group members, can be influenced through variations in the salience of social categorization.

References

Abelson, R. P. (1995). *Statistics as principled argument.* Hillsdale, NJ: Erlbaum.

Aristotle (1963). *Organon graece* (Vol. 2). Dubuque, IA: W. C. Brown Reprint Library. (Original work published 1846)

Bacon, F. (1960). *The new organon.* Indianapolis, IN: Bobbs-Merrill. (Original work published 1620)

Brewer, M. B., & Kramer, R. M. (1985). The psychology of intergroup attitudes and behavior. *Annual Review of Psychology, 36,* 219–243.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *MMPI-2 manual for administration and scoring.* Minneapolis: University of Minnesota Press.

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin, 54,* 297–312.

Cantor, N., & Mischel, W. (1979). Prototypes in person perception. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 12, pp. 3–52). New York: Academic Press.

Clement, R. W., & Krueger, J. (1995, June). *Social projection in a double minimal group paradigm.* Poster session presented at the Seventh Annual Convention of the American Psychological Society, New York.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Dawes, R. M. (1989). Statistical criteria for a truly false consensus effect. *Journal of Experimental Social Psychology, 25,* 1–17.

Edwards, W. (1982). Conservatism in human information processing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 359–369). Cambridge, England: Cambridge University Press.

Einhorn, H. J., & Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological Review, 92,* 433–461.

Forer, B. R. (1949). The fallacy of personal validation: A classroom demonstration of gullibility. *Journal of Abnormal and Social Psychology, 44,* 118–123.

Gaertner, S. L., Mann, J., Murrell, A., & Dovidio, J. F. (1989). Reducing intergroup bias: The benefits of recategorization. *Journal of Personality and Social Psychology, 57,* 239–249.

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics.* Hillsdale, NJ: Erlbaum.

Hays, W. L. (1988). *Statistics* (4th ed.). New York: Holt, Rinehart and Winston.

Hoch, S. J. (1987). Perceived consensus and predictive accuracy: The pros and cons of projection. *Journal of Personality and Social Psychology, 53,* 221–234.

Howard, J. W., & Rothbart, M. (1980). Social categorization and memory for in-group and out-group behavior. *Journal of Personality and Social Psychology, 38,* 301–210.

Hume, D. (1955). *An inquiry concerning human understanding.* Indianapolis, IN: Bobbs-Merrill. (Original work published 1748)

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge, England: Cambridge University Press.

Keppel, G., Saufley, W. H., Jr., & Tokunaga, H. (1992). *Introduction to design and analysis* (2nd ed.). New York: Freeman.

Krueger, J. (1992). On the overestimation of between-group differences. *European Review of Social Psychology, 3,* 31–56.

Krueger, J., & Clement, R. W. (1994a). Memory-based judgments about multiple categories: A revision and extension of Tajfel's accentuation theory. *Journal of Personality and Social Psychology, 67,* 35–47.

Krueger, J., & Clement, R. W. (1994b). The truly false consensus effect: An ineradicable and egocentric bias in social perception. *Journal of Personality and Social Psychology, 67,* 596–610.

Krueger, J., & Zeiger, J. S. (1993). Social categorization and the truly false consensus effect. *Journal of Personality and Social Psychology, 65,* 670–680.

Lingle, J. H., Altom, M. W., & Medin, D. L. (1984). Of cabbages and kings: Assessing the extendibility of natural object concept models to social things. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 71–118). Hillsdale, NJ: Erlbaum.

Mackie, D. M., Allison, S. T., Worth, L. T., & Asuncion, A. G. (1992). The generalization of outcome-biased counter-stereotypic inferences. *Journal of Experimental Social Psychology, 28,* 43–64.

Markman, E. M. (1989). *Categorization and naming in children.* Cambridge, MA: MIT Press.

McNemar, Q. (1962). *Psychological statistics* (3rd ed.). New York: Wiley.

Mill, J. S. (1974). *A system of logic ratiocinative and inductive.* Toronto, Canada: University of Toronto Press. (Original work published 1843)

Mullen, B., Dovidio, J. F., Johnson, C., & Copper, C. (1992). In-group–out-group differences in social projection. *Journal of Experimental Social Psychology, 28,* 422–440.

Myers, I. B. (1962). *The Myers-Briggs Type Indicator Manual.* Princeton, NJ: Educational Testing Service.

Nisbett, R., & Ross, L. (1980). *Human inference.* Englewood Cliffs, NJ: Prentice Hall.

Oakes, P., & Turner, J. (1990). Is limited information processing capacity the cause of social stereotyping? *European Review of Social Psychology, 1,* 111–136.

Rabbie, J. M., & Horwitz, M. (1969). Arousal of ingroup–outgroup bias by a chance win or loss. *Journal of Personality and Social Psychology, 13,* 269–277.

Reichenbach, H. (1951). *The rise of scientific philosophy.* Berkeley: University of California Press.
Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior, 14,* 665–681.
Rorer, L. G. (1991). Some myths about psychology. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology* (Vol. 1, pp. 61–90). Minneapolis: University of Minnesota Press.
Rosch, E. H., & Lloyd, B. B. (1978). *Cognition and categorization.* Hillsdale, NJ: Erlbaum.
Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13,* 279–301.
Rothbart, M. (1981). Memory processes and social beliefs. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 145–182). Hillsdale, NJ: Erlbaum.
Rothbart, M., & Lewis, S. B. (1988). Inferring category attributes from exemplar attributes: Geometric shapes and social categories. *Journal of Personality and Social Psychology, 55,* 861–872.
Rothbart, M., & Taylor, M. (1992). Category labels and social reality: Do we view social categories as natural kinds? In G. R. Semin & K. Fiedler (Eds.), *Language, interaction and social cognition* (pp. 11–36). London: Sage.
Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.
Simon, B. (1993). On the asymmetry in the cognitive construal of ingroup and outgroup: A model of egocentric social categorization. *European Journal of Social Psychology, 23,* 131–148.
Smith, E. E. (1991). Categorization. In D. N. Osherson & E. E. Smith (Eds.), *An invitation to cognitive science* (Vol. 3, pp. 33–53). Cambridge, MA: MIT Press.
Snyder, C. R., & Shenkel, R. J. (1976). Effects of "favorability," modality, and relevance on acceptance of general personality interpretations prior to and after receiving diagnostic feedback. *Journal of Consulting and Clinical Psychology, 44,* 34–41.
Sokal, R. R., & Rohlf, F. J. (1969). *Biometry.* San Francisco: Freeman.
Spears, R., & Manstead, S. R. (1990). Consensus estimation in social context. *European Review of Social Psychology, 1,* 81–109.
Sundberg, N. D. (1955). The acceptance of "fake" versus "bona fide" personality test interpretations. *Journal of Abnormal and Social Psychology, 50,* 145–147.
SuperPaint 3.0 [computer software]. (1993). San Diego, CA: Aldus Corporation.
Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American, 223,* 596–102.
Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76,* 105–110.
Weber, R. & Crocker, J. (1983). Cognitive processes in the revision of stereotypic beliefs. *Journal of Personality and Social Psychology, 45,* 961–977.

Appendix A

Minnesota Multiphasic Personality Inventory (MMPI-2) Statements Used as Stimulus Materials in Experiment 2

1. I seldom worry about my health
2. At times I have very much wanted to leave home.
3. I think I would like the kind of work that a forest ranger does.
4. I think most people would lie to get ahead.
5. I enjoy detective or mystery stories.
6. I like to go to parties or other affairs where there is lots of loud fun.
7. I have very few headaches.
8. I have never done anything dangerous for the thrill of it.
9. I often think, "I wish I were a child again."
10. I do not worry about catching diseases.

(Appendix B follows on next page)

Appendix B

Statements Used as Stimulus Materials in Experiment 3

Table B1
Statements Used for the Assessment of Projection (No Mean Change)

Statement	Time 1 *M*	Time 1 *SD*	Time 2 *M*	Time 2 *SD*	*p*	*d*
1. During a speech, I am self-conscious about my posture, diction, and ability to make sense.	6.9	3.3	6.7	3.6	.357	.06
2. Before giving a speech, it is more important to clarify what one wants to say than what the audience is ready to hear.	4.6	3.9	4.7	5.2	.720	.02
3. Late at night, I would rather watch Letterman than Leno.	7.2	3.9	7.0	6.2	.272	.05
4. The art of rhetoric has not made much progress since the time of Aristotle.	4.5	2.9	4.2	2.9	.229	.09
5. I need to cut down on my coffee consumption.	3.6	8.7	3.2	8.1	.282	.04
6. I have difficulty starting to do things.	4.7	6.6	4.6	6.8	.742	.01
7. I am neither gaining nor losing weight.	5.9	7.5	5.8	7.1	.937	.00
8. I think nearly everyone would tell a lie to keep out of trouble.	5.0	6.5	5.1	5.1	.659	.02

Table B2
Discarded Statements (Mean Change)

Statement	Time 1 *M*	Time 1 *SD*	Time 2 *M*	Time 2 *SD*	*p*	*d*
1. I feel nervous before and during a public speech or presentation.	7.0	3.4	4.9	3.6	.000	.60
2. Every time I am about to give a speech, I look forward to it.	4.4	5.2	5.2	3.4	.000	.20
3. Usually, I hope that no one will ask me questions after a presentation.	3.6	3.8	3.1	3.0	.017	.16
4. I am familiar with the main principles of rhetoric.	4.3	4.3	6.7	3.9	.000	.59
5. When I attend a public speech, I scrutinize the speaker's presentational skills.	6.4	4.1	7.5	2.8	.001	.30
6. At least once, I felt terrific and satisfied during a speech because I knew I had nailed it.	6.7	6.1	7.5	4.1	.021	.16
7. When I give a speech, I have a tendency to avoid making eye contact with the audience.	3.8	6.3	2.7	3.6	.000	.23
8. I feel uncomfortable with the thought that I should take up a lot of space when I speak.	4.4	5.4	3.4	4.0	.000	.21
9. When I speak it is most important to me that people like me.	5.3	5.3	4.7	5.0	.013	.12
10. When I speak I sense the resonance in my diaphragm.	2.6	2.7	3.7	3.8	.000	.30
11. The audience can always tell when I become anxious or lose focus.	5.7	4.4	4.2	4.2	.000	.34
12. Usually I am able to get my ideas across to the audience.	6.3	2.8	6.9	2.5	.003	.24

Note. Means are based on the ratings of the communication students (experimental condition and group control).

Received March 21, 1995
Revision received August 14, 1995
Accepted August 15, 1995 ■