Journal of Personality and Social Psychology
1994, Vol. 67, No. 4, 596–610

0022-3514/94/$3.00

The Truly False Consensus Effect: An Ineradicable and Egocentric Bias in Social Perception

Joachim Krueger and Russell W. Clement

Consensus bias is the overuse of self-related knowledge in estimating the prevalence of attributes in a population. The bias seems statistically appropriate (Dawes, 1989), but according to the egocentrism hypothesis, it merely mimics normative inductive reasoning. In Experiment 1, Ss made population estimates for agreement with each of 40 personality inventory statements. Even Ss who had been educated about the consensus bias, or had received feedback about actual consensus,or both showed the bias. In Experiment 2, Ss attributed bias to another person, but their own consensus estimates were more affected by their own response to the item than by the other person's response. In Experiment 3, there was bias even in the presence of unanimous information from 20 randomly chosen others. In all 3 experiments, Ss continued to show consensus bias despite the availability of other statistical information.

In a study on student attitudes, Katz and Allport (1931) noticed that the more students admitted they had cheated on an exam, the more they expected that other students cheated too. Since then, more than a hundred studies have documented a systematic relationship between people's perceptions of their own characteristics and their estimates of the percentage of people in the population who share those characteristics. Early investigators assumed that the cause of this relationship is that people irrationally project their own characteristics onto others. Much research effort was dedicated to the examination of the psychological causes of projection (Holmes, 1968). Ross, Greene, and House (1977) considered projection to be a consensus bias (i.e., the "false-consensus effect") and introduced it to the attribution and decision-making literature. These authors reinforced the idea that consensus bias is irrational. This argument has two parts. First, a person's own response to a judgment item is a single-case sample. To the extent that other social information is available, the self-related single-case sample provides little information and should be ignored in the inference process. Second, if consensus estimates vary with the person's own response, at least some of the estimates must be incorrect. If raters ignored their own responses, there would be no differences between the mean estimates of people with different responses.

The assumption that consensus bias stems from flawed reasoning has been challenged. Dawes (1989) reexamined the data obtained by Ross, Greene, and House (1977) and argued from a Bayesian perspective that subjects were correct in considering their own behavioral choices common in the population. According to this analysis, even a sample of 1 should have substantial effects on percentage estimates. Therefore, it is conceivable that subjects in research on consensus bias intuitively understand the logic of statistical induction and perform accordingly. Empirically, however, the observed consensus bias tends to be larger than is statistically appropriate (Krueger & Zeiger, 1993). This finding raises the possibility that statistical (i.e., Bayesian) reasoning may not play any role in consensus estimates at all. We review methods of separating statistically appropriate consensus effects from true bias and then develop the egocentrism hypothesis. According to this hypothesis, consensus bias does not result from Bayesian thinking, but from less analytical cognitive processes. We then report three experiments in which subjects are presented with various kinds of information that should reduce bias if integrated in a statistically appropriate way.

When the False Consensus Effect Is Truly False

The standard test of bias is whether the mean consensus estimate provided by people who endorse an item is greater than the mean estimate provided by those who do not endorse the item. If the means differ, at least one of them is inaccurate. Inaccurate estimates do not necessarily imply flawed reasoning (Einhorn, 1986). Endorsers do not have the same sample information that nonendorsers have. At least one piece of information, the estimators' own response to the judgment item, is different. If people followed statistical principles of induction, they should honor all available sample information, and hence, endorsers should make higher consensus estimates than nonendorsers (Dawes, 1989, 1990; Hoch, 1987; Krueger & Zeiger, 1993).

Suppose a woman enjoys Bergman movies, whereas her fiancé does not. She also believes that Bergman movies are more popular than he does. Similarly, if she draws a blue chip from an urn of unknown contents, whereas he does not draw a sample, her estimate of the percentage of blue chips should be higher

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

We thank Tami Bryan and Landon Reid for help with data collection and Hal Arkes, Russ Church, Robyn Dawes, Jacob Ham, Oliver John, Neil Macrae, and Jill Portman for comments on a draft of this article.

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

than his. If they are each unaware of the information the other person has, these differences in estimation may be justified by inductive reasoning alone. The optimal difference between their estimates can be derived from Bayes's rule if all a priori probabilities (e.g., of enjoying Bergman movies or drawing a blue chip) are known. In generic induction tasks, this condition can be controlled experimentally. In social prediction, however, the prior probabilities may not be known.[1] Moreover, the computation of optimal posterior probabilities (to be presented later in this article; see also Dawes, 1989) is sufficiently complex to cast doubt on the idea that the average social perceiver can perform the necessary calculations consciously and reliably.

If the analysis is extended to multiple items, it is easier to separate true bias from appropriate induction. Across items, various within-subjects correlations can be computed. Because one is more likely to espouse popular than unpopular attitudes, a person's attitudes (or item endorsements) tend to be correlated with actual consensus (i.e., the percentage of people who endorse the item). The correlation between actual consensus and a person's endorsements ($r_{act,end}$) expresses *self-validity.* It is well-known that consensus estimates tend to be correlated with endorsements. This correlation expresses *simple projection* ($r_{est,end}$). Because endorsements tend to be valid, people who engage in simple projection are more likely to achieve *correlational accuracy* (i.e., the correlation between estimated and actual consensus [$r_{est,act}$]) than people who do not (Hoch, 1987).

To understand that in principle, consensus bias (i.e., simple projection) is justified, it is crucial to realize that for the majority of raters, endorsements are positively correlated with actual consensus. The average person's self-validity is positive ($r_{act,end} > 0$) regardless of what the percentages of actual consensus are and regardless of whether item endorsements are independent or correlated. The exception to the rule is when actual consensus is the same for all items. In that case, the denominator of the correlation formula is 0 and the coefficient is not defined. Consider a numerical example in which endorsements of four items are uncorrelated and the size of the majority on each item is 70%. If 70% of movie-goers like Actor A, 70% like Actor B, 30% like Actor C, and 30% like Actor D, 65% of the self-validity coefficients are positive.[2] The most probable specific pattern of endorsements is the one that is perfectly correlated with actual consensus (i.e., liking A and B and disliking C and D): $P(r_{est,act} = 1.0) = P(A) \times P(B) \times P(1 - C) \times P(1 - D) = .7^4 = .24$. Self-validity is negative for only 8%. The least probable specific pattern of endorsements is the one that is perfectly inversely correlated with actual consensus (i.e., disliking A and B and liking C and D): $P(r_{est,\,act} = -1.0) = P(1 - A) \times P(1 - B) \times P(C) \times P(D) = .3^4 = .008$. Self-validity is zero for 18%, and the correlation is not defined for 9% (when all four actors are either liked or disliked).[3]

Now suppose that raters are aware of the validity of their item endorsements. That is, they rightly assume that most of their endorsements reflect majority positions. Unless self-validity is perfect, the raters also hold at least one minority position. If they do not know on which items they are in the minority, their optimal strategy is to assume that they are in the majority on all items. Someone who likes Actors A, B, and C but dislikes Actor D has positive self-validity ($r_{act,end} = .58$) and may reasonably assume that A, B, and C are more popular than D. Possible consensus estimates are 80% for each of Actors A, B, and C and 20% for Actor D. These estimates would be quite accurate ($r_{est,act} = .58$) and simple projection would be perfect ($r_{est,end} = 1.0$). If the estimates were 60% for Actors A, B, and C and 40% for Actor D, correlational accuracy and simple projection would be the same. Note, however, that although both sets of estimates are consistent with the optimal inference strategy, the rater systematically overprojects in the first case and underprojects in the second. The measure that is sensitive to the difference between over- and underprojection is the correlation between endorsements and the differences between estimated and actual consensus (10, 10, 50, and −10 in the first case [$r_{diff,end} = .66$] and −10, −10, 30, and 10 in the second case [$r_{diff,end} = -.17$]). People overproject if they believe that relative to actual consensus their preferences are more common than their alternatives. When positive, this correlation expresses a "truly false consensus effect" (hereafter, TFCE; see also Krueger & Zeiger, 1993). The TFCE indexes the irrational component of consensus bias.

Can Consensus Bias Be Eliminated?

If consensus bias reflected only statistically appropriate thinking, simple projection ($r_{est,end}$) and correlational accuracy ($r_{est,act}$), but not the TFCE ($r_{diff,end}$) should be greater than 0. In a first test of the TFCE, however, all three correlations were significant (Krueger & Zeiger, 1993). Subjects were presented with statements from the revised Minnesota Multiphasic Personality Inventory (MMPI-2; e.g., "I like to flirt"; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) and estimated the percentage of people who would endorse each item. Relative to the actual percentages, subjects' estimates were higher for those items that they themselves endorsed than for those that they did not endorse. This TFCE occurred for judgments about the general population and gender in-groups but not for judgments about out-groups. Two interpretations of these results are possible. Subjects may have deliberately followed the appropriate inductive strategy of generalizing from themselves to groups they belonged to, but in the process generalized too much. Alternatively, the task of making population estimates may trigger fairly automatic and egocentric inferences that others who belong to the same group are similar to the rater. To test whether

[1] Prior probabilities are the likelihoods of specific outcomes (e.g., that 25% of the chips in an urn are blue) before any sample information has been gathered.

[2] Five of the 16 possible endorsement patterns are positively correlated with the actual consensus. If item endorsements are uncorrelated with each other, the probability that a given pattern occurs is the product of the probabilities of each item response. The sum of the probabilities of the five patterns that are positively correlated with the actual consensus across items is 65%.

[3] The size of the majority with positive self-validity is moderated by the homogeneity of the population and the intercorrelations between item endorsements. The more the average actual consensus deviates from 50%, the more homogeneous is the population and the more likely it is that a person's endorsements will represent the actual consensus. Furthermore, if item endorsements are intercorrelated, a person who is in the majority on one item is more likely to be in the majority on another item. Even if endorsements are negatively correlated across items, however, most correlations of self-validity are positive.

consensus bias results from superficial egocentric reasoning rather than statistical analysis, it is necessary to provide the rater with additional statistical information. If people reason egocentrically, they will continue to base consensus estimates largely on their own responses. If they reason statistically, they will weigh the additional information appropriately and consensus bias will be diminished.

Increasing the amount of relevant information appears to be the prime recipe for improving judgment (Fischhoff, 1982). Relevant information can come in various forms. Standard debiasing techniques involve the use of instructional material that explains the nature of a bias to subjects before they engage in the judgment task. Another method is to provide accuracy feedback after each judgment. In Experiment 1, we examined whether such direct debiasing techniques can diminish the TFCE ($r_{\mathrm{diff,end}}$) while leaving the optimal strategy of simple projection ($r_{\mathrm{est,end}}$) intact. A more indirect form of information is the presentation of responses made by other subjects. If the response of the observed other varies independently of the subject's response, and if subjects then have the opportunity to revise their own estimates, consensus bias should be reduced. In Experiment 2, we tested whether subjects attribute consensus bias to others and whether taking the other's perspective improves their own subsequent estimates. Third, social prediction can be viewed as a special case of generic induction because population characteristics are probabilistically inferred from sample information. To understand the peculiarities in the use of sample information in self-based social prediction, we juxtaposed social and generic induction in Experiment 3.

Across experiments, the egocentrism hypothesis holds that self-related information is treated as superior to other sample information. Therefore, subjects will show consensus bias even when debiasing techniques are used (Experiment 1), will underuse other-related information in social prediction (Experiment 2), and will neglect sample information in generic induction (Experiment 3).

Experiment 1: Debiasing

The design of Experiment 1 followed the within-subjects correlational approach, which permitted the assessment of individual differences in the degree of bias and accuracy (Krueger & Zeiger, 1993). The *TFCE* is the correlation between the difference between estimated and actual consensus and item endorsements ($r_{\mathrm{diff,end}}$). *Simple projection* is the correlation between consensus estimates and endorsements ($r_{\mathrm{est,end}}$). *Self-validity* is the correlation between actual consensus and endorsements ($r_{\mathrm{act,end}}$). *Correlational accuracy* is the correlation between estimated and actual consensus ($r_{\mathrm{est,act}}$). The *standard consensus bias* occurs when people who endorse an item give higher consensus estimates than people who do not endorse the item. Finally, *mean-level accuracy* is the absolute average within-subjects difference between estimated and actual consensus.

To test the robustness of consensus bias, we used three strategies. First, the experiment was designed to minimize consensus bias in any condition. Earlier work has shown that consensus bias is relatively small (a) when the target population is highly inclusive, (b) when the number of items judged is large, and (c) when estimates follow endorsements (Mullen et al., 1985). Thus, subjects were asked to estimate consensus in the general (i.e., inclusive) adult population for many (40) items after they had made their own endorsements (order). If the egocentrism hypothesis is correct, consensus bias will appear even under these restrictive conditions.

Second, two debiasing techniques (feedback and education) were manipulated experimentally. Feedback consisted of the display of the actual consensus of each item immediately after the subject had made the estimate. The availability of accuracy information provided an opportunity to detect over- and underestimation and gradually calibrate judgment. Education is the direct approach of explaining the biasing role of self-knowledge in population estimates and exhorting subjects not to succumb to it (Fischhoff, 1975). Giving or withholding education or feedback resulted in a two-factorial between-subjects design. Debiasing should be greatest when subjects have been informed about the nature of the TFCE and obtain on-line accuracy information. If, however, the egocentrism hypothesis is correct, self-knowledge will bias consensus estimates regardless of the availability of feedback or education.

Third, it is possible that consensus bias, in part, results from people's tendency to ascribe positive rather than negative attributes to both themselves and others (Sherman, Chassin, Presson, & Agostinelli, 1984). To control this potential confound, items were also rated on social desirability (SD). The within-subjects correlations between endorsements and social desirability ratings (*self-image* = $r_{\mathrm{SD,end}}$) and between consensus estimates and social desirability ratings (*other-image* = $r_{\mathrm{SD,est}}$) were expected to be positive. According to the egocentrism hypothesis, simple projection and the TFCE will be significant even when the variance in social desirability ratings is partialed out.

Method

Subjects. One hundred twenty-two (62% women) Brown University undergraduates served as subjects in exchange for credit for an introductory psychology course. They participated in groups of 1–8.

Procedures and design. On entering the laboratory, subjects were told that the experiment was a study on "social judgment." They were seated in individual cubicles equipped with Macintosh IIci computers, and instructions for the separate components of the experimental session appeared on the screen. Over the course of 1 hr, subjects were presented with 40 statements from the MMPI-2 (Butcher et al., 1989) three times. Each time, statements appeared individually and remained on the screen until the subjects responded.

After the presentation of each of the 40 items, subjects did or did not endorse the statement by clicking a box labeled *agree* or *disagree.* After completing the 40 judgments, subjects worked on an unrelated task for 5–10 min. For the second presentation, they were instructed to rate how socially desirable it is to agree with an item. This rating was made for each item on a scale ranging from *socially undesirable* (1) to *socially desirable* (9). Then, subjects worked again on an unrelated task for 5–10 min. When the items were presented for the third time, subjects were instructed to "enter the percentages between 0 and 100 that best reflect your belief about the proportion of people who would agree with each statement." In the baseline condition there were no further instructions. In the education condition, subjects received the following additional information:

> Please note that previous research indicates that these types of estimates are affected by the rater's own agreement or disagreement

with the statement. When people agree with a statement, they usually give a high estimate relative to the actual percentage of agreement in the population. In contrast, when people disagree, they usually give a low estimate relative to the actual percentage. With this information in mind, please try to be as accurate with your estimates as possible.

In the feedback condition, subjects were told that after each estimate they would also "see the actual percentage of agreement." Subjects typed in their estimates while the statement was displayed. In sum, the design had four conditions: a baseline condition and three debiasing conditions. In the debiasing conditions, subjects received either education, feedback, or both.

The 40 items are displayed in Table 1 along with the rates of actual consensus as reported in the MMPI-2 manual. Criteria for item selection were similar to those reported in Krueger and Zeiger (1993). Statements suggesting personality pathology and statements with extreme actual consensus (above 80% or below 20%) were not included.

Results

Between-subjects analyses. For a standard test of consensus bias, estimates were averaged within items and across experimental conditions and separately for endorsers and nonendorsers. The data in Table 1 show that for each of the 40 statements, the mean consensus estimate was higher among subjects who agreed with it than among subjects who disagreed with it. With a Bonferroni-adjusted alpha ($p < .001$ for two-tailed t tests), 19 of these comparisons (48%) were significant. To compare endorsers' and nonendorsers' estimates in each condition, the means of the estimates were averaged across items. This was done separately for the means obtained from endorsers and nonendorsers and separately for each condition. Unweighted means were used in this analysis and results are displayed in Table 2.

A 2 (endorsements) × 2 (education) × 2 (feedback) between-cases analysis of variance (ANOVA) was performed on the means of the consensus estimates. In this analysis, items rather than subjects were the cases. There was only a significant effect of endorsement, $F(1, 312) = 81.0, p < .001$. The absence of any effect involving conditions of debiasing supported the egocentrism hypothesis (all other Fs < 1).

Within-subjects analyses. Table 3 presents the average within-subjects correlations (resulting from r-to-Z-to-r transformations; see McNemar, 1962) for each of the four conditions.

The average Z scores were tested against 0 by means of two-tailed t tests, and then the effects of the experimental manipulations were tested by 2 (education) × 2 (feedback) between-subjects ANOVAs. Only effects reliable at the .01 level were considered significant. Simple projection was significant ($r_{\text{est,end}} = .35$, $p < .001$), but it was reduced by neither education, $F(1, 120) = 1.3$, nor feedback, $F(1, 120) = 3.7, p > .05$, nor the combination of the two ($F < 1$). The average TFCE was significant ($r_{\text{diff,end}} = .16, p < .001$) and did not diminish when education or feedback was given, all $F\text{s}(1, 120) < 1.8$. In each condition, a large proportion of subjects had a positive correlation ($r_{\text{diff,end}} > 0$). The proportions were 67%, 70%, 90%, and 71% in the baseline, the education-only, the feedback-only, and the education-and-feedback conditions, respectively. As predicted by the egocentrism hypothesis, the TFCE and simple projection resisted the combined forces of two debiasing techniques. The smaller size of TFCE relative to simple projection demonstrates that $r_{\text{diff,end}}$ is the more conservative measure of bias.[4]

Not surprisingly, subjects preferred to endorse desirable over undesirable statements. These positive self-images ($r_{\text{SD,end}} = .24$, $p < .001$) did not vary across conditions (all Fs < 1.8). Also as expected, consensus estimates tended to be higher for desirable than for undesirable statements ($r_{\text{SD,est}} = .15, p < .001$). Curiously, these other-images were more positive in the conditions with feedback than without, $F(1, 120) = 10.9, p < .001$. Could these social desirability effects have spuriously inflated consensus bias? To test this possibility, simple projection and TFCE were computed again as partial correlations, controlling for the covariance with SD. Both correlations remained unchanged ($r_{\text{est,end} \times \text{SD}} = .33, p < .001$) and ($r_{\text{diff,end} \times \text{SD}} = .17, p < .001$), and their size did not vary across conditions (all Fs < 1.8).

Accuracy. Before we turn to analyses concerning judgmental accuracy, recall that a person's endorsements tend to be informative about actual consensus. Self-validity was evidenced by the positive correlation between actual consensus and endorsements ($r_{\text{act,end}} = .18, p < .001$), and this correlation did not vary across conditions (all Fs < 1). Overall, correlational accuracy was modest but significant ($r_{\text{est,act}} = .07, p < .001$), and it was greater when feedback was available than when it was not available, $F(1, 120) = 14.1, p < .001$. The effect of education and the interaction did not reach the chosen level of significance (Fs $= 6.1$ and 4.7, respectively, $p > .01$). According to the Bayesian analysis of consensus bias, some degree of bias is necessary to maximize accuracy. To test whether correlational accuracy would have been smaller if subjects had not shown any bias, the correlations between estimated and actual consensus were computed again while item endorsements were statistically controlled. The grand mean of these partial correlations was essentially 0 ($r_{\text{est,act} \times \text{end}} = .01$). It was significantly smaller than the mean of the zero-order correlations, $F(1, 120) = 62.8, p < .001$, and this effect did not vary across conditions, $F(1, 120) < 1$. That is, correlational accuracy would have been entirely absent had subjects not projected.

Variations in the size of the differences between estimated and actual consensus have little effect on the correlational indices of bias and accuracy. Is it possible that the debiasing techniques increased mean-level accuracy while preserving the correlational biases and only modestly improved correlational accuracy? The means of the absolute differences between estimated and actual consensus were computed for each subject. Mean differences were larger in the baseline condition ($M = 22.39$) and the education-without-feedback condition ($M = 22.50$) than in the two conditions including feedback (Ms $= 19.27$ and 18.25 with and without education), $F(1, 120) = 31.0, p < .001$. An analysis of the mean standard deviations of estimates yielded similar results. When feedback was provided, the mean variability of the estimates (Ms $= 19.16$ and 17.97 with and without education) was smaller than when no feedback was provided (Ms $= 20.34$ and 22.86 with and without education), $F(1, 120) = 16.2, p < .001$, thus approaching the degree of homoge-

[4] Not surprisingly, the difference scores were positively correlated with consensus estimates ($r_{\text{est,diff}} = .76, p < .001$), but they were negatively correlated with SD ($r_{\text{SD,diff}} = -.13, p < .001$).

Table 1
Actual and Estimated Consensus of 40 MMPI Items

Items	MMPI-2	Endorsers	Nonendorsers	$p<$
1. I sweat very easily even on cool days.	21	44.54	29.26	.001
2. My conduct is largely controlled by the behavior of those around me.	28	60.15	49.36	.004
3. My hardest battles are with myself.	73	62.80	46.22	.002
4. I like to be with a crowd who play jokes on one another.	24	55.96	40.99	.001
5. I have very few fears compared to my friends.	54	50.71	36.61	.001
6. I like poetry.	62	55.81	47.80	.025
7. I am easily awakened by noise.	48	54.20	53.74	.875
8. I never indulged in any unusual sex practices.	70	55.35	50.96	.325
9. I seldom worry about my health.	64	44.67	34.40	.007
10. I enjoy reading love stories.	47	53.49	47.12	.052
11. I like to let people know where I stand on things.	75	66.01	61.87	.197
12. I certainly feel useless at times.	36	64.12	40.71	.001
13. At times I have very much wanted to leave home.	37	67.11	48.92	.001
14. It does not bother me that I am not better looking.	60	41.16	28.12	.001
15. I think I would like the kind of work that a forest ranger does.	51	46.10	27.00	.001
16. In school I found it very hard to talk in front of the class.	56	57.80	51.47	.058
17. I am neither gaining nor losing weight.	65	49.07	39.70	.010
18. I would like to be a singer.	43	56.50	39.71	.001
19. I used to keep a diary.	40	55.60	50.00	.112
20. I enjoy a race or a game more when I bet on it.	30	60.59	44.93	.001
21. I think most people would lie to get ahead.	48	66.12	48.36	.001
22. I worry over money and business.	54	66.75	62.13	.103
23. I work under a great deal of tension.	37	64.18	59.24	.102
24. I have no fear of spiders.	52	50.74	37.09	.001
25. I am embarrassed by dirty stories.	29	48.43	44.96	.287
26. I enjoy detective or mystery stories.	67	55.20	46.41	.003
27. I am a very sociable person.	71	65.16	59.17	.036
28. I like to read newspaper articles on crime.	45	58.98	43.28	.001
29. Criticism or scolding hurts me terribly.	47	60.44	44.40	.001
30. I like to go to parties or other affairs where there is lots of loud fun.	42	65.25	59.20	.050
31. I have very few headaches.	80	53.18	45.81	.018
32. I like collecting flowers or growing house plants	61	47.80	43.14	.105
33. My sex life is satisfactory.	74	53.12	42.46	.005
34. I have never done anything dangerous for the thrill of it.	39	47.52	35.87	.006
35. I do not mind being made fun of.	36	42.00	26.57	.001
36. I like dramatics.	63	54.42	45.61	.001
37. I often think, "I wish I were a child again."	22	69.24	53.44	.001
38. I am so touchy on some subjects that I can't talk about them.	25	52.55	39.15	.001
39. My eyesight is as good as it has been for years.	57	52.27	38.77	.001
40. I do not worry about catching diseases.	64	48.00	33.18	.001

Note. MMPI-2 = revised Minnesota Multiphasic Personality Inventory.

Table 2
Unweighted Means of Population Estimates Across Items

	Education			
	Yes		No	
	Feedback		Feedback	
Endorsement	Yes	No	Yes	No
Yes	54.76	54.95	54.81	57.69
No	45.71	42.98	43.93	45.02

neity in the actual consensus data (M = 16.13). No other effects were significant.

Individual differences. The within-subjects correlational approach provided an opportunity to examine individual differences in the degree of consensus bias. The relationship between self-validity and projection was particularly interesting. The higher the self-validity ($r_{act,end}$), the more representative the person is of the population. To be accurate, a person with high self-validity should project more than a person with low self-validity. The data showed, however, that subjects did not know the extent of the validity of their own endorsements. Correlational accuracy was low because subjects of varying self-validity projected to the same extent. Across subjects, self-validity and simple projection were uncorrelated (r = .003), a finding that has an important consequence for the TFCE. If people project regardless of their self-validity, those whose endorsements are representative (i.e., valid) of the group will show the smallest TFCE. By contrast, people who endorse uncommon attributes that they believe to be common and who do not endorse common attributes that they believe to be rare, will produce difference scores that are highly correlated with their endorsements.

Table 3
Mean Within-Subjects Correlations as a Function of Education and Feedback

	Education			
	Yes		No	
	Feedback		Feedback	
Correlations and variables	Yes	No	Yes	No
Zero-order correlations				
TFCE ($r_{end,diff}$)	.22**	.15*	.16**	.12
Simple projection ($r_{end,est}$)	.41**	.31**	.35**	.31**
Self-validity ($r_{act,end}$)	.16**	.20**	.18**	.18**
Accuracy ($r_{est,act}$)	.14**	.09	.12*	−.08
Self-image ($r_{end,SD}$)	.25**	.29**	.21**	.21**
Person-positivity ($r_{est,SD}$)	.24**	.17*	.20**	−.03
Partial correlations				
TFCE × SD ($r_{end,diff \times SD}$)	.24**	.18**	.17**	.19**
Simple projection × SD ($r_{end,est \times SD}$)	.38**	.29**	.31**	.31**
Accuracy × endorsements ($r_{est,act \times end}$)	.07	.01	.06	−.12

Note. TFCE = truly false consensus effect; SD = social desirability.
* $p < .01$. ** $p < .001$.

Indeed, the degree of the TFCE was negatively correlated with self-validity across subjects ($r = -.58$, $p < .05$).

Discussion

Experiment 1 documented the robustness of consensus bias in three ways. First, neither education about the nature of consensus bias, nor on-line feedback about actual consensus, nor the combination of the two reduced simple projection or the TFCE. Between-subjects and within-subjects analyses yielded convergent evidence for the failure of debiasing. Second, consensus bias was not a byproduct of people's tendency to endorse socially desirable statements (positive self-image) and their belief that people in general endorse desirable statements (positive other-image). Third, all experimental conditions shared features that, according to previous research, should minimize bias: The target population was highly inclusive, the number of judgment items was large, and subjects made endorsements before they estimated consensus. The exception to the pattern of persistent bias was a modest improvement in correlational and mean-level accuracy in response to feedback about actual consensus. Subjects learned that their estimates were too extreme and in the course of the experiment made more regressive and more accurate estimates, although any individual piece of feedback was uncorrelated with the actual consensus of the following statement.

Arkes (1991) suggested that direct debiasing methods such as education or feedback improve inferences only when biases are "strategy-based," that is, when judges misconstrue the problem or are too lazy to think through the task. The absence of debiasing and the modest improvements in accuracy in Experiment 1 suggested that consensus bias is not strategy-based. A different type of bias is "association-based," resulting from simple, symmetrical, and nonstatistical connections between cognitive elements (Arkes, 1991). According to this view, the self-descriptiveness of an item is automatically associated with high consensus estimates without requiring explicit statistical reasoning.

The direct debiasing techniques relied on multiple items and different groups of subjects responding to different instructions or information. Indirect methods offer an alternative route, involving single items and within-subjects tests. The key to indirect debiasing is to induce decision makers to consider counterfactual events (e.g., choices they had not made) and to estimate their likelihood. Thinking about an explanation for an event that did not happen (Ross, Lepper, Strack, & Steinmetz, 1977) or simply imagining the event affects probability estimates (Sherman, Cialdini, Schwartzman, & Reynolds, 1985). To the extent that the estimated probability of a counterfactual event increases, the estimated probability of the actual event may decrease and thus be less biased than if the counterfactual had not been considered. This "consider-the-opposite" strategy has been used to reduce overconfidence and hindsight biases (Arkes, Faust, Guilmette, & Hart, 1988; Lord, Lepper, & Preston, 1984).

In studies on consensus bias, the provision of information about the behavior of others has had mixed results. Either subjects ignored sample information (Hansen & Donoghue, 1977) or they took it into account only under certain conditions, for example, when their self-esteem was not threatened (Sherman, Presson, & Chassin, 1984) or when the others were particularly representative of the population (Zuckerman, Mann, & Bernieri, 1982). Goethals (1986) found that consensus bias disappeared when the presented samples reflected actual consensus in the population. So far, no study has examined the effect of other-related information within subjects. Experiment 2 was designed to do this. Subjects estimated consensus on a single item and learned about another person's endorsement. They then tried to infer the other person's consensus estimate. Finally, they had the opportunity to revise their own estimates. The egocentrism hypothesis was that subjects would persist with their initial consensus estimate even when the other disagreed with their choice and even when they realized that the other would make a divergent consensus estimate. In other words, egocentric projection may be sufficiently strong to survive the challenge from an indirect debiasing technique.

Experiment 2: Self–Other Differences

Earlier work has demonstrated that subjects attribute consensus bias to others. In a simulation of the Ross, Greene, and House (1977) sandwich board study, subjects believed that those participants in the sandwich board study who complied with the experimenter's request estimated the percentage of compliance to be higher than did those participants who declined to comply (Krueger & Zeiger, 1993, Experiment 4). The size of the attributed consensus effect was virtually identical to the size of the actual consensus effect. Can the finding that consensus bias is attributed to others be used to extract diverging estimates from the same subject on the same item? Perhaps subjects base consensus estimates on their own choices and at the same time concede that another person, whose choices are different, will provide different estimates consistent with those choices.

To use an example from the public domain, suppose a president nominates a friend for a high office, and he or she is optimistic that most political decision makers support the candidacy. If the president assumes, however, that support for the candidate is not unanimous, he or she may expect that the opposition is also sure of victory. This realization involves the insight that consensus estimates depend on the position of the judging person. Different estimates made by self and other cannot both be correct. Only if both estimators were unaware of the other's position could both estimates be optimal without necessarily being accurate or identical (Dawes, 1989). One's own choice, even if merely hypothetical, is accessible and practically irrepressible. Thus, information about another person's choice is the second observation in a sample of 2 and should have considerable impact on estimates.

According to the egocentrism hypothesis, consensus bias will persist when the estimator knows that another person's position on an item is different from his or her own. Statistically, the source of an observation is irrelevant, as long as it is randomly sampled. When sample size increases from 1 (self-related information) to 2 (self- plus other-related information), self-related information is no more privileged than other-related information. Thus, if they were statistically derived, consensus estimates should be moderated when information becomes available about someone who disagrees with the rater on the item being judged. Returning to the example, suppose the president learns that a specific committee member opposes the proposed appointment. Combining his or her own preference with the additional diverging observation, the president should now make a more cautious estimate about the support of the protégé, especially if the president realizes that the opposing committee member's estimate is biased against the candidate. To do this, the president could average his or her own prior estimate and the estimate he or she attributes to the opponent.

The robustness of the TFCE, as observed in Experiment 1, suggests that subjects, unlike the thoughtful but hypothetical president, will not average their own consensus estimates with the estimates they attribute to a disagreeing other. Although they may attribute biased consensus estimates to the other, they may assume that their own estimates are impartial and closer to the truth or that their own projection is more justified because they consider themselves more typical or representative of the population. Attributing projection to others while overlooking one's own projection is egocentric. To summarize, three predictions were derived from the egocentrism hypothesis. First, the standard within-item and between-subjects consensus effect should replicate. Second, consensus effects should be attributed to others, and the size of this effect should be as large as the original self-related consensus effects. Third and most important, subjects should fail to revise their estimates after exposure to the choice of a randomly drawn other and after attributing consensus bias to that other.

Method

Subjects. Ninety-seven (63% women) undergraduate students at Brown University participated in exchange for extra credit in an introductory psychology course or a small payment ($5). They were tested in groups of 1–8.

Procedures and design. Experiment 2 was conducted in the same setting and with the same cover story as Experiment 1. The experiment had three phases, separated by unrelated tasks. Subjects made standard self-related consensus estimates, estimates attributed to another person, and follow-up self-related estimates. The order of the first self-related estimates and the other-related estimates was varied. About half the subjects made self-related estimates first, followed by other-related estimates, whereas the other half made these estimates in reverse order. All subjects concluded by making self-related estimates again. In describing the procedures, we will follow the first of these two orders.

In Phase 1, subjects read a statement about a personal characteristic ("Criticism or scolding hurts me terribly"). They indicated whether they agreed or disagreed with it by clicking the appropriate box (labeled *agree* or *disagree*). On a separate screen, they then entered their consensus estimates. The specific MMPI item was selected because it had produced a strong consensus bias in Experiment 1 (*M*s = 60.44% and 44.40% for endorsers and nonendorsers, respectively), and its actual consensus lay close to one in two (47%).

In Phase 2, instructions read:

> You will now be presented with a statement and whether another individual agrees or disagrees with it. The information concerning the other individual will be drawn at random from a data base of subjects who have previously participated in this experiment.

Subjects clicked a box labeled *"Access Data Base,"* whereupon a flickering cursor and disk activity, which lasted for several seconds, created the impression of a random access operation. In fact, random assignment to condition at the beginning of the experiment had determined whether subjects learned about another person who agreed or

Table 4
Mean Follow-up Percentage Estimates for Agree Response

	Self	
Other	Yes	No
Yes	71.14	47.22
No	58.89	40.42

disagreed with the statement. After reading the endorsement that the other person had ostensibly made, subjects received the following instructions:

> We would now like you to estimate this other person's belief about the percentage of people who agree with this statement. Enter the number from 0 to 100 that corresponds to your best guess.

Phase 3 was a repetition of Phase 1. Subjects were again asked to supply their own (self-related) consensus estimates. This repeated measure provided an opportunity to revise earlier estimates in light of the encountered other-related information. The three dependent variables (initial self-related estimates, attributed estimates to the other, and follow-up self-related estimates) were collected in a design with three between-subjects variables: own endorsement (yes vs. no), other's endorsement (yes vs. no), and order of initial self-related and other-related judgments.

Results

Separate 2 (own endorsement) × 2 (other's endorsement) × 2 (order) between-subjects ANOVAs were conducted for the three dependent variables. Order did not significantly affect any of the dependent variables and is omitted in the presentation of the results. As in Experiment 1, effects were considered significant if they were reliable at the level of $p < .01$.

Initial estimates made for self and other. The standard consensus bias emerged as an effect of endorsement by self. Those who agreed with the statement, "Criticism or scolding hurts me terribly" believed that more people endorse this statement ($M = 65.53$) than did subjects who did not agree with the statement ($M = 45.39$), $F(1, 96) = 29.3$, $p < .001$. Endorsements by others had no effect on self-related estimates, $F(1, 96) = 1.9$. As expected, subjects attributed consensus bias to others. Those who learned that the other had agreed with the statement believed that the person would make a higher estimate ($M = 65.96$) than did those who learned that the other had disagreed ($M = 48.46$), $F(1, 96) = 23.1$, $p < .001$. The size of this attributed consensus effect was almost identical to the size of the self-based consensus effect. Subjects' own positions had no effect on the attributed estimates, $F(1, 96) < 1$.

Follow-up estimates made for self. As predicted by the egocentrism hypothesis, follow-up estimates (Phase 3) were virtually identical to the means of the initial estimates. Results are shown in Table 4.

Consensus estimates were higher among subjects who agreed with the statement ($M = 65.02$) than among those who disagreed ($M = 43.82$), $F(1, 96) = 35.3$, $p < .001$. Furthermore, subjects who had learned that the other student had agreed tended to give higher estimates ($M = 59.20$) than those who had learned that the other student had disagreed ($M = 49.66$). The size of this effect was less than half (difference = 9.52) of the effect of the subjects' own endorsement (difference = 21.20) and did not reach the selected level of significance, $F(1, 96) = 6.2$, $p > .01$. No other effects were significant.

Differences in weight given to one's own and to others' endorsements were most evident in the conditions where the endorsements of self and other were discrepant. Statistically, it should not matter whether oneself or somebody else had judged the item. Contradictory information obtained from the sample of 2 should cancel each other out.[5] If, however, subjects assumed egocentrically that their own endorsements were more informative, consensus bias should persist. The data in Table 4 show that, when averaged across order conditions, consensus estimates were higher among subjects who agreed with the statement while the other disagreed ($M = 58.89$) than among subjects who disagreed while the other agreed ($M = 47.22$).

Discussion

The standard within-item and between-subjects consensus bias was replicated, and at the same time, subjects attributed consensus bias to others. Most important, subjects showed little inclination to incorporate the other's position in their consensus estimates. In the revised estimates, the weight assigned to their own position was more than twice that of the weight given to the other's position. When we discovered the attribution of consensus bias to others (Krueger & Zeiger, 1993), it seemed that people knew that they project and expect others to do the same. The failure of the present subjects to adjust their estimates after attributing projection to others suggests instead that their own projection remained undetected.

The egocentric pattern of projection, paired with the attribution of projection to others and the maintenance of the belief that one's own estimates are more accurate, fits Holmes's (1968) concept of "similarity projection." Similarity projection is the projection "onto other individuals [of] traits *identical* to those which [the perceiver] possesses but the possession of which he is *not aware*" (p. 259, emphasis in the original). Ironically, the findings in Experiment 2 indicated just this in the domain of projection itself. Subjects may not have realized that they projected but believed that others did.

The sample-size heuristic and the law of large numbers. When subjects gave less weight to other-related than to self-related information, they violated the statistical law of large numbers. This law describes a monotonic relationship between sample size and the reliability of parameter estimates. Normally, a sample of $n + 1$ observations is a better estimate of population characteristics than a sample of n observations. In generic in-

[5] In a sample of 2, one *agree* and one *disagree* response cancel each other out only when the prior probability of agreement is 50%. If the prior probability were higher, the improbable *disagree* response would carry greater weight than the probable *agree* response and reduce the posterior probability of agreement. In the present case, however, the assumption that the prior probability of agreement with the statement is close to 50% is justified because (a) actual consensus in the national sample was 47% (Butcher et al., 1989) and 61% among participating subjects, and (b) the unweighted average of the initial self-related estimates made by agreers and disagreers was 55.46%.

duction, where self-related information is unrelated to the estimation task, people realize that percentage estimates should increase with increasing samples of unanimous information. In the well-known "shreeble study," subjects inferred the characteristics of an exotic species of bird from sample data. The larger the all-blue sample of shreebles was, the higher were the percentage estimates of blue shreebles in the species (Nisbett, Krantz, Jepson, & Kunda, 1983). This sample-size heuristic was used for a variety of categories (e.g., obese tribespeople, electricity-conducting metals, and marbles in urns) and regardless of whether the sample data uniformly indicated the presence or the absence of the rated feature (Krueger, 1994; Peterson, Schneider, & Miller, 1965).

In social prediction, too, intuitive estimates are sensitive to sample size, as long as self-related information is excluded. Rothbart (1981) described "bookkeeping" as one way of forming and changing social stereotypes. This strategy of mental arithmetic involves the storage and integration of information about observed group members in memory. Beliefs about the characteristics of the group undergo gradual adjustments when discrepant information becomes available. Empirically, the bookkeeping model describes social and nonsocial category learning quite well (Krueger, 1991; Rothbart & Lewis, 1988).

Predictive conservatism. Despite subjects' awareness of the law of large numbers in generic and other-related social prediction, intuitive induction is not good enough. When samples are small, predictions are usually too conservative. People underestimate the degree to which diagnostic information changes base rates (Edwards, 1982). Predictive conservatism is evident in experiments where subjects draw chips from an urn and estimate the probability that most chips in the urn are of the sampled color (Peterson et al., 1965). Suppose there are two urns, B and R, one with a ratio of blue to red chips of 60%:40%, and the other with a ratio of 40%:60%. A priori, each urn is equally likely to be presented (i.e., P[B] = .5). After the random draw of a blue chip, the probability that the urn predominantly contains blue chips (P[B/blue]) changes from .5 to .6. This result follows from Bayes's rule that the posterior probability of having selected the urn with mostly blue chips given that the sample chip was blue (P[B/blue]) is equal to the prior probability of selecting an urn of mostly blue chips (P[B]) multiplied with the likelihood ratio (P[blue/B]/P[blue]). The likelihood ratio is the probability of drawing a blue chip given the urn that predominantly contains blue chips divided by the overall a priori probability of drawing a blue chip. Hence, $.5 \times .6 / .5 = .6$. Typically, subjects fail to recognize the consequences of a single-item sample and continue to believe that the probability that they are drawing from the urn that predominantly contains blue chips is .5.

The present findings and previous studies on probabilistic inference suggest a dissociation between intuitions about generic and other-related social induction on the one hand and self-related social predictions on the other hand. In generic induction, people follow statistical reasoning by making larger changes in their predictions as sample size increases, but the size of the adjustments is insufficient. In contrast, people do not treat self-related information as an ordinary sample of 1, but as qualitatively distinct information of high diagnostic value, whose impact on population predictions (i.e., consensus estimates) is unmitigated by other available social information. Experiment 3 used a revision-of-probability procedure to directly compare generic and self-based social prediction.

Experiment 3: Social Versus Nonsocial Prediction

The social prediction task consisted of a simulation of the sandwich board study (Ross, Greene, & House, 1977). Subjects estimated the percentage of students who would comply with the experimenter's request to help in a persuasion study. The generic induction task involved estimating the percentage of blue chips in an urn. In both parts, samples provided uniform evidence (all sampled students complied; all drawn chips were blue). Sample size increased from 0 to 1 to 3 and to 20.

The first hypothesis was that people would use a sample-size heuristic in generic induction and other-related social prediction. That is, percentage estimates (i.e., consensus estimates in the social part) should increase with sample size. The second hypothesis was that predictions would show the conservatism bias. That is, revisions of probability estimates should be insufficient regardless of sample size. Specifically, people will tend not to recognize that the first piece of sample data is the most informative and that it entails greater optimal statistical change from prior to posterior probability than any additional piece of evidence of the same type. For example, in normative prediction, drawing another blue chip or encountering yet another absent-minded professor yields successively smaller change. The third hypothesis, egocentrism, was that when subjects' own choices are taken into account, the standard consensus bias would return. When no other social information is available, self-related consensus bias may make people seem to conform to Bayes's rule when in fact they are making optimal judgments for the wrong (egocentric) reasons. The biased nature of egocentric projection should become apparent with increasing sample size. Consensus estimates were expected to go up, but the magnitude of the adjustments would be insufficient and the gap between agreers and disagreers would not close as much as it should.

Bayes's Rule for Multiple Prior Probabilities

Dawes (1989) suggested that the false consensus effect may not be false because its typical size is similar to the statistically normative change from prior to posterior probabilities in generic induction. The normative change can be precisely calculated in the chips-and-urns paradigm because the assumptions entering the task can be stated explicitly. If there are 100 chips in an urn, but the ratio of reds to blues is unknown, there are 101 binomial hypotheses. In the simplest case, each possible percentage of blue chips is equally likely a priori (i.e., p_{urn} = .0099). Aggregating across hypotheses, the prior probability of drawing a blue chip is $p_{blue} = .5$, which is the sum of the products of each prior probability and the probability of drawing a blue chip from each specific urn (i.e., $\Sigma(p_{urn} \times p_{blue/urn})$. Given these assumptions, the normative change in the probability of blue chips after the draw of one blue chip (i.e., going from p_{blue} to $p_{blue/blue}$) can be calculated in a two-step procedure.

First, the probability of each of the 101 possible distributions needs to be revised. Because it is more likely that the blue chip was drawn from a predominantly blue urn than from a predominantly red urn, the probabilities of the former go up and the probabilities of the latter go down. The probability of the all-red urn becomes nil. Consider, for example, the posterior probability of the 80% blue urn as an example. According to Bayes's rule:

$$\text{P(blue/urn)} = \text{P(urn)} \times \frac{\text{P(blue/urn)}}{\text{P(urn)}}.$$

That is, .0158 = .0099 × .8 /.5. Second, the posterior probability of each distribution is multiplied with the likelihood of drawing a blue chip given that distribution (P[blue/urn]). Then, the products are summed across the 101 distributions so that P(blue/blue) = Σ(P[urn/blue] × P[blue/urn]) = .67. When the population is large enough (roughly $N > 100$), the Bayesian analysis can be reduced to the formula P(blue/blue) = (k + 1)/(n + 2), where k is the number of "successes" (e.g., blue chips drawn) and n is the sample size (Dawes, 1989). A quick calculation shows the negative acceleration of Bayesian induction. When $k = n = 1$, $p = .67$; when $k = n = 2$, $p = .75$; when $k = n = 3$, $p = .8$; and so forth. When the population is large, it is irrelevant whether a small sample is replaced.[6]

Social prediction differs from generic induction because the prior probabilities are implicit. Still, many social prediction tasks approach the psychological conditions of generic statistical problems. Social judges often have limited sample information in combination with a wide range of plausible prior probabilities. It would be unreasonable to ask social perceivers to be cognizant of the prior probabilities of all possible percentage distributions, especially when the prior probabilities are not uniform. Moreover, it is unlikely that perceivers without formal training master Bayes's rule of properly combining prior probabilities in calculating posterior probabilities. We do maintain, however, that to understand probabilistic intuition, it is necessary to compare intuitive with normative prediction. In Experiment 3, about one third of the subjects were presented with the chips-in-urn problem described above, and the other two thirds were presented with a problem of social prediction that approaches the chips-in-urn problem in terms of prior uncertainty.

Method

Subjects. A total of 319 undergraduate students (71% women) volunteered as subjects for this experiment. Some were enrolled at the University of Rhode Island and others at Brown University. Of these participants, 222 responded to a questionnaire on social prediction and 97 completed a questionnaire on generic induction.

Procedures and design. The social prediction task consisted of a simulation of a classic experiment in which subjects were asked to help in a study on persuasion by walking around the Stanford campus wearing a sandwich board with the words *Eat at Joe's* or *Repent* (for a detailed description of the instructions, see Ross, Greene, & House, 1977). After indicating whether they were willing to help, the Stanford subjects estimated the percentage of students who would comply. Compliant subjects estimated compliance to be more prevalent than did noncompliant subjects.

In the present experiment, subjects read about the procedures of the Stanford study. About half were presented with the *Repent* version and half with the *Eat at Joe's* version. They were then asked, "What percentage of students do you think agreed to wear the sign?" After writing down their estimate, subjects turned to the next page, which presented information about the putative choices of samples of Stanford students.

> Now suppose you happened to meet one of the participants in the Stanford study by chance. This student tells you that he agreed to participate in the attitude study. Again please estimate the number of students who agreed to wear the sign.

After making the second estimate, subjects received information about 3 and finally about 20 Stanford subjects who ostensibly had all agreed to the request. After each stage of sample information, subjects reentered an estimate. Subjects also responded to the following query: "If you had been a participant in the Stanford study, would you have agreed to walk around with the sandwich board?" About half of the subjects entered their own behavioral choices before making the consensus estimates. The other half entered their choices at the end of the experiment, just before they were debriefed and dismissed.

Subjects who participated in the generic induction part of the experiment read the following instructions:

> This questionnaire is part of a study on human judgment. One type of judgment is called *induction.* People make inductive inferences whenever they estimate the characteristics of a large group of objects based on their knowledge of "samples" of observations. In this questionnaire you will find several hypothetical scenarios. Please read these scenarios carefully. You will then be asked to make probabilistic estimates (percentages). Imagine your task is to estimate the color of objects in an urn. Let's say the objects are chips. You know that there are *100* chips in the urn, and that the only possible colors that chips can be are blue or red. Although you do not know the exact composition of the urn, you know that 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.* Given the above assumptions, what is your best guess of the percentage of blue chips in the urn?

After making a percentage estimate, subjects were asked to imagine they had drawn at random 1 blue chip from the urn. They estimated again, and the procedure was repeated with 3 and 20 chips, thus keeping the sample sizes comparable with those in the social prediction part of the experiment.

Results

Results are displayed in Figure 1. The top curve shows optimal Bayesian predictions in generic induction (i.e., P = [k + 1]/[n + 2]), followed in descending order by estimates in the chips-and-urn task and social consensus estimates made by subjects who would have agreed or disagreed to carry the sandwich board.

Social prediction. Percentage estimates of compliance were analyzed in a 2 (endorsement: yes vs. no) × 2 (sex) × 2 (condition: *Eat at Joe's* vs. *Repent*) × 2 (order: estimates first vs. own choice first) × 4 (sample size: 0, 1, 3, or 20) ANOVA with repeated measures on the last variable. As expected, subjects used a sample-size heuristic by increasing their estimates with increasing samples of compliant group members, $F(3, 630)$ =

[6] In the present work, uniform prior probabilities were chosen because they represent the most cautious set of hypotheses in a situation of complete uncertainty. Bayesian posterior probabilities can also be calculated when the prior probabilities are not uniform. Nonuniform prior probabilities can vary considerably, as long as their sum is equal to 1.0.

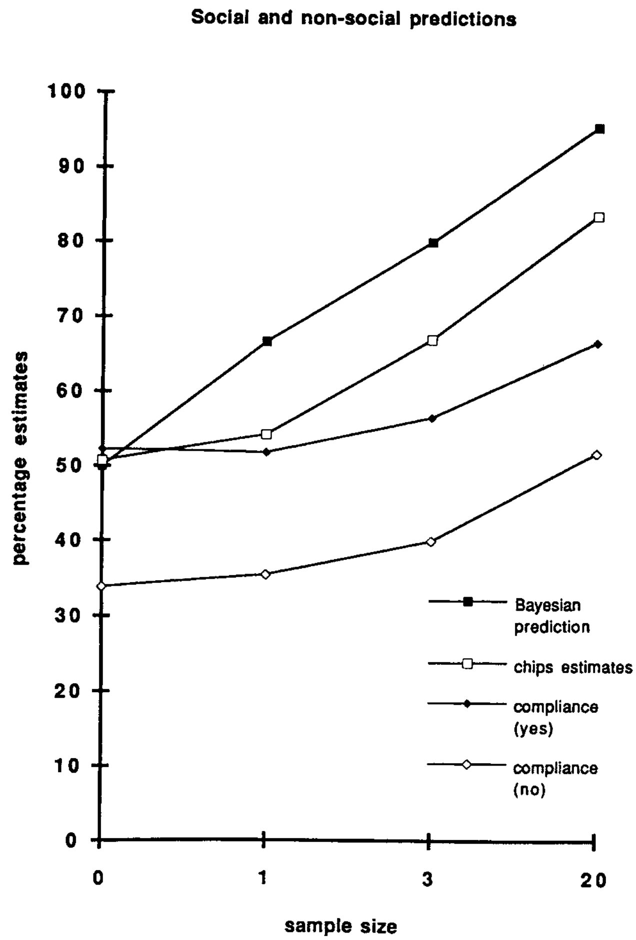

Figure 1. Bayesian predictions compared with nonsocial and social percentage estimates as a function of sample size. *Note.* The unequal differences between sample sizes (1 − 0 = 1; 3 − 1 = 2; 20 − 3 = 17) make the Bayesian predictions appear more linear than they are.

105.4, $p < .001$. However, as predicted by the conservatism hypothesis, the size of this effect was too small. Introducing information about one other person did not change the initial estimates ($n = 0$), $t(218) = 1.4$, *ns*. Samples of 3 and 20 others led to successive increases, $t(218) = 7.9$ and 12.3, all $ps < .001$. Yet, even the increases from estimates based on samples of 3 to estimates based on a sample of 20 unanimously acting others (10.09 and 11.8 for agreers and disagreers, respectively) were merely half the size of the initial ($n = 0$) difference between agreers and disagreers (19.76).

As predicted by the egocentrism hypothesis, subjects who would have agreed to carry the sandwich board believed that compliance was more prevalent than did subjects who would not have complied (see Figure 1), $F(1, 210) = 28.4$, $p < .001$. Most important, the difference between the estimates made by agreeing and disagreeing subjects remained constant across increasing sample information. The lack of an interaction between sample size and subjects' own endorsements documented this phenomenon, $F(3, 630) < 1$. Even when 20 randomly sampled Stanford students were said to comply, subjects used their own preference as a guidepost to infer the preferences of others. Furthermore, consensus estimates did not depend on whether subjects had indicated their own behavioral choice before or after being exposed to the sample (all *F*s involving order < 1).

The order of making one's own behavioral choice (however hypothetical) and estimating the behavior of others was irrelevant, suggesting that the effect of subjects' own choices was particularly robust. From a statistical perspective, sample information should have affected the choices themselves. If one learns that all of the 20 randomly sampled others responded in a certain way, it is more likely that one would act in the same way compared with a situation without sample information. This should be true particularly when the critical behavior is novel and does not involve prior experience. Contrary to this reasoning, the probability of agreeing to comply was not significantly greater when sample information about compliance preceded own choices ($p = .43$) than when it followed own choices ($p = .31$), $\chi^2(1, N = 222) = 3.4$, $p > .05$.

Generic induction. In estimating the proportion of red chips in an urn, subjects followed the heuristic that large samples are more informative than small samples (see Figure 1). A one-way repeated-measures ANOVA with four levels of sample size (0, 1, 3, and 20) was significant, $F(3, 288) = 171.6$, $p < .001$, and so were all three two-tailed paired t tests (df = 96) comparing estimates for adjacent levels of sample size (all $ps < .001$). Although estimates increased with sample size, the increments were too small. On all three levels of sample size, estimates fell significantly below the optimum Bayesian probability (all $ps < .001$). Only when no sample information was given, subjects estimated the optimal percentage of blue chips (50%) with sufficient accuracy, $t(96) = .6$, *ns*.

From a Bayesian perspective, the first data point in a sample is the most informative, requiring the largest adjustment from prior to posterior probabilities. Additional adjustments become successively smaller with increasing sample size. Subjects' predictive conservatism revealed a counternormative philosophy. Most subjects seemed to discount the first data point sampled as uninformative. Both the mode (75% of subjects) and the median estimate remained at 50% after the first draw.

Social versus nonsocial prediction. Both social and nonsocial predictions were too conservative, and the use of the first available data point was particularly insufficient. Inspection of the data in Figure 1 indicates that predictive conservatism was even stronger in social than in nonsocial prediction. To test whether revisions of probability estimates were significantly greater in generic than in social prediction, estimates based on a given sample size were subtracted from estimates based on the next larger sample. This analysis showed that initial adjustments (from $n = 0$ to $n = 1$) did not differ significantly for social and generic prediction, $t(184) < 1$. Subsequent adjustments, however, were larger in generic than in social prediction, to $n = 3$: $t(138) = 5.3$, $p < .001$, and to $n = 20$: $t(197) = 2.1$, $p < .04$.

Retrospective conservatism. Predicting conservatively is failing to realize how similar the population is to the observed sample. It follows that subjects would have to retrospectively underestimate the likelihood that the observed sample would have occurred in the first place. Recall that the posterior probability of blue is $21/22 = .9545$ if all the 20 draws were blue. Supposing that $p = .9545$ is known to be true, the probability of observing 20 successes in 20 draws is the optimal posterior probability of success to the power of the sample size ($p = .9545^{20} = .40$). In the generic induction task, the estimated posterior probability of blue was $p = .836$. Given this belief, the retrospective probability of having drawn a run of 20 blue would have to be $p = .836^{20} = .0278$. Even a modest conservatism bias ($.9545 - .836 = .1185$) implies a greatly reduced probability of obtaining the specific sample that produced the prediction. In social prediction, estimates of the posterior probability of compliance implied even lower retrospective probabilities of finding unanimous behavior in a sample of 20 students ($p = .667^{20} = .000304$ for compliant subjects and $p = .518^{20} = .00000193$ for noncompliant subjects). That is, the insufficient use of sample information in social prediction makes the very observations that led to the conservative estimates look like a statistical oddity. This retrospective analysis of the probability of obtaining the sample in the first place illustrates the implausibility of conservative predictions about the population.

Discussion

The data supported the three hypotheses. First, in both social and generic prediction, subjects used the sample-size heuristic, gradually increasing consensus estimates as more unanimous information became available. Second, estimates were conservative. Subjects underestimated the diagnostic value of randomly sampled data. Third, estimates showed egocentric consensus bias rather than conservatism when subjects had only their own position as sample information to rely on. The absence of conservatism in self-based prediction need not reflect adequate Bayesian inference. Instead, it is possible the egocentrism bias cancels out the conservatism bias. The persistence of consensus bias (a difference of 14.89%) even in the presence of sample information for about 20 others suggested that subjects did not view their own choices as "just another piece of data." Consider how small the normative impact of a single piece of data is in the generic induction task. The difference in the posterior probability of getting 19 or 20 successes out of 20 draws is 4.55% (20/22 − 21/22).

The larger the sample, the smaller is the value of one's own position in population prediction. Egocentrism may supply ways, however, of discounting the predictive power of large samples of other-related information. Because random samples are rarely perfectly reliable, many have mistakenly concluded that such samples are altogether uninformative. It is this very randomness, however, that ensures a measure of predictive validity (Dawes, 1988). One version of dismissing random sampling is to point out cases where different random samples have failed to yield identical results. Before the 1992 presidential election, President Bush continued to believe he enjoyed the support of the majority of the public, although most polls showed otherwise. "There's something crazy about the polling . . . they can't all be right, so some have to be nutty" (President Bush on Larry King Live, October 30, 1992). To the follow-up suggestion "When *you* get closer [the polls] are not crazy, though" he replied "Well, maybe when *you* get closer" (emphasis added). When different samples are available, the egocentric choice is to believe the data that confirm one's own projection.

The role of egocentrism in consensus estimates has not been fully realized because people's sensitivity to sample size and the role of conservatism in generic induction have been challenged by the view that people follow a mistaken "law of small numbers" and overgeneralize in any prediction domain (Tversky & Kahneman, 1971). Dawes (1988) concluded that "a single instance is a poor basis for generalization [but] nevertheless, such generalization occurs—often with great ease" (pp. 97–98). Similarly, Nisbett and Ross (1980) emphasized people's "willingness to make strong inferences based on small amounts of data" (p. 81). However, these authors conceded that the insensitivity to sample size occurred only when "consideration of sample size has been pitted against the potent representativeness heuristic, and in each instance the former has been vanquished by the latter" (p. 81). In the present study, use of the sample-size heuristic and conservatism reemerged when the confound between sample size and representativeness was removed. Subjects overused information from the single-case sample only when the information was self-related.

The combination of egocentric consensus estimates and predictive conservatism with other-related information place a burden on social relationships and hamper the revision of social stereotypes. People are more surprised about the actions of others than about their own, especially when others behave differently from how they, the observers, would. In daily life, the incredulous "I-can't-believe-you-did-that" attitude is inevitable if one believes that (a) others generally share one's preferences (egocentric consensus bias) and that (b) if they do not, they must belong to a highly atypical minority (retrospective conservatism). This self-serving pattern of inference resists disconfirmation. Not even exposure to uniformly behaving others effectively combats the impression that the observed behavior is rare. These findings suggest that social beliefs (e.g., stereotypes) are resistant to change because exemplar-based information, drawn from observing group members, yields insufficient updates of group-related beliefs.

General Discussion

In three experiments, consensus bias survived debiasing efforts virtually unchanged. These results support the egocentrism hypothesis and challenge the Bayesian perspective. According to the Bayesian perspective, rational subjects would have taken additional information (i.e., feedback or other-related information) into account to reduce bias. This did not occur. The term *egocentrism* stresses the nonstatistical reasoning underlying consensus bias, and it aptly suggests rigidity of judgment and a sense of special value of self. In its current form, however, the egocentrism hypothesis says little about the mechanisms underlying biased judgment. Can the existing process-oriented explanations of consensus bias account for the present data?

Process-Oriented Explanations

Cognitive explanations stress the potential of selective exposure, selective attention, and selective memory for self-related attributes to sway consensus estimates (e.g., Ross, Greene, & House, 1977). Selective information processing is a statistically inadequate strategy. Feedback, in its various forms, should have reduced bias by bringing other relevant information into view. Because bias persisted, we suspect that cognitive selectivity effects are not the main source of consensus bias. Could one argue that self-related information is more salient than other-related information? In the present experiments, other-related information was not salient. Information came in a numerical format, without the actual presentation of the person. The salience argument is not convincing, however, because it fails to account for the results in Experiment 2. If, for lack of salience, subjects failed to consider the positions of hypothetical others in their own consensus estimates, they should also have been unable to attribute consensus bias to these others. To test the salience hypothesis rigorously, future research will have to examine whether people persist in ignoring other people's positions when those others are well-known (i.e., salient) individuals rather than anonymous students as in Experiment 2.

Motivational explanations emphasize the self-protective or self-enhancing function of consensus bias (e.g., Sherman, Presson, & Chassin, 1984). Self-protective or self-enhancing processes are usually assessed by varying the type of the item or the state of the perceiver (e.g., by presenting a threat to the subject's self-esteem). Interestingly, both consensus bias and the putative false uniqueness effect have been traced to the need to feel good about oneself. Some research indicates that self-protection can enhance consensus bias (Sherman, Presson, & Chassin, 1984), but this does not mean that the minimally sufficient source for consensus bias is motivational. In the present research, social desirability effects did not contribute to consensus bias, but the possibility remains that egocentric projection involves a general motivation to see others as similar to oneself regardless of the desirability of the attribute.

Primitive Cognition as a Cause of Egocentrism

None of the two process-oriented theories explain all the data, but each of them has received partial support in the past. In concluding this article, we discuss the assumption of causation underlying both theories and suggest an extension of the cognitive approach that may provide a satisfactory model for the presented evidence.

Causation. Research on consensus bias has tacitly assumed that subjects' item endorsements cause high or low consensus estimates rather than vice versa. Interestingly, however, most of the evidence is correlational. Unless endorsements are manipulated directly, comparisons between the mean estimates provided by endorsers and by nonendorsers merely test the correlation between these subject groups and consensus estimates. Similarly, the within-subjects analyses assess correlations between endorsements and estimates across items. Because correlations do not express causation, one might as well entertain the possibility that making high consensus estimates causes people to agree with items and making low estimates causes them to disagree. Such inferences may seem absurd to consensus researchers who take stable preferences for granted but seem reasonable to students of conformity. Some people make patently inaccurate perceptual judgments when confronted with the judgments of a unanimous but mistaken majority (Asch, 1956); purchasing decisions are easily swayed by "social proof" indicating that certain products are popular (Cialdini, 1984), and responses to personality inventory questions depend in part on perceptions of what the socially normative responses are (Paulhus, 1984).

The egocentrism hypothesis shares the assumption of causation that is implicit in all consensus research, and Experiment 3 provided tentative empirical support for this idea. If consensus estimates had caused subjects' own endorsements, subjects who had learned about the unanimous behavior of a large sample of students should have aligned their own behavioral choices with the majority. However, the rate of compliance among these subjects was not greater than among subjects who had made their own decisions before they were exposed to the sample information. Possibly, the present procedures did not involve direct conformity pressures, and it remains to be seen whether such pressures may produce a reversal of the commonly accepted route of causation in consensus bias.

To establish the causal role of own endorsements in consensus bias, research will have to move from correlational to experimental designs. Some investigators have begun to explore the effects of within-subjects changes of position on consensus estimates. McCauley, Durham, Copley, and Johnson (1985) found that patients who had undergone successful kidney transplants estimated the success rate of such transplants to be higher than patients whose transplants were not successful or patients on a waiting list. Similarly, Agostinelli, Sherman, Presson, and Chassin (1992) found consensus bias after arbitrary feedback following a problem-solving task (Sherman, Presson, & Chassin, 1984). To be fully conclusive, complete experimental designs will involve pretests of endorsements and estimates, followed by an experimental manipulation of either the endorsements or the consensus estimates, followed by posttests of endorsements and estimates. If the prevailing theory of self-related causation is correct, consensus estimates will increase for those subjects whose responses to the items changed from *disagree* to *agree*, and estimates will decrease for those subjects whose responses changed from *agree* to *disagree*. Moreover, after manipulations of subjects' consensus estimates, their item endorsements should not change.

Primitive cognition. If we tentatively accept the idea that people's choices and preferences play a causal role in shaping their perceptions of population characteristics, we can consider three aspects of the cognitive approach as possible explanations of projective egocentrism. First, making adequate inductive inferences requires an understanding of sampling procedures. Throughout this article we maintain that people should regard themselves as single-case samples randomly drawn from a population because this is what they are from a statistical point of view. Lacking individuating knowledge about a given subject, a particular subject is as representative or unrepresentative as the next subject. From the subject's perspective, however, the self is not "randomly drawn." Others may be considered random samples because there is less individuating information associated with others than with the self. Others can be ignored or discounted as atypical of the population. The self may not appear as a sample because more individuating information is available and because self-related information predates any sampling activity. Therefore, the person may conclude, however erroneously, that self-related information is particularly informative about population characteristics.

The second aspect of a cognitive view is concerned with order. In the typical consensus experiment, the source of social information (self or other) is confounded with the order of availability. A rudimentary (self-related) affective response to an item may come to mind easily, even when it was not solicited (Zajonc, 1980), whereas other-related information takes more time to be transmitted. To test the idea that self-related information is particularly powerful data because it predates information sampled from others, future research will need to control the order of presentation more tightly. So far, experiments have started with the presentation of the target items. Even if other-related information is presented next, subjects may have already made their own covert response. That is, self-related information always enjoys the advantages of primacy. To circumvent the confound of order, the other's response to the item could be presented before the item itself. If consensus bias persists, the egocentrism hypothesis would be strengthened. Research in generic induction has shown that data favoring one hypothesis are used insufficiently when they are preceded by data favoring an alternative hypothesis (Peterson & DuCharme, 1967).

Embedded in this version of the primacy effect is the idea that access to self-related information may be more automatic than access to other-related information. Conventional cognitive explanations of consensus bias emphasize the effects of conscious and deliberate thought. Selective exposure to similar others and attention to and retrieval of their attributes suggest controlled, if biased, information processing. A considerable amount of research has shown that many mental activities occur fast, automatically, and even outside of awareness (Uleman & Bargh, 1989). The present data are consistent with the view that for most people there is a fundamental association between the self and the social norm, an association operating independently from controlled statistical reasoning. Hence, the idea that "most people are like me" may be spontaneous. If such automatic associations exist, future research will have to determine its developmental sources. Perhaps egocentric population inferences are developmental vestiges of the infantile belief that all others are like us.

References

Agostinelli, G., Sherman, S. J., Presson, C. C., & Chassin, L. (1992). Self-protection and self-enhancement: Biases in estimates of population prevalence. *Personality and Social Psychology Bulletin, 18,* 631–642.

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin, 110,* 486–498.

Arkes, H. R., Faust, D., Guilmette, T. J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology, 73,* 305–307.

Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs, 70* (Whole No. 416).

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.

Cialdini, R. B. (1984). *Influence: The new psychology of modern persuasion.* New York: Quill.

Dawes, R. M. (1988). *Rational choice in an uncertain world.* New York: Harcourt Brace Jovanovich.

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

Dawes, R. M. (1990). The potential nonfalsity of the false consensus effect. In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn* (pp. 179–199). Chicago: University of Chicago Press.

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. (1986). Accepting error to make less error. *Journal of Personality Assessment, 50,* 387–395.

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 288–299.

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422–444). Cambridge, England: Cambridge University Press.

Goethals, G. R. (1986). Fabricating and ignoring social reality: Self-serving estimates of consensus. In J. Olson, C. P. Herman, & M. P. Zanna (Eds.), *Ontario symposium on personality and social psychology: Vol. 4. Relative deprivation and social comparison* (pp. 135–157). Hillsdale, NJ: Erlbaum.

Hansen, R. D., & Donoghue, J. M. (1977). The power of consensus: Information derived from one's own and others' behavior. *Journal of Personality and Social Psychology, 35,* 294–302.

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

Holmes, D. S. (1968). Dimensions of projection. *Psychological Bulletin, 69,* 248–268.

Katz, D., & Allport, F. (1931). *Students' attitudes.* Syracuse, NY: Craftsman Press.

Krueger, J. (1991). Accentuation effects and illusory change in exemplar-based category learning. *European Journal of Social Psychology, 21,* 37–48.

Krueger, J. (1994). *Generalization from instances.* Manuscript in preparation.

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

Larry King Live. (1992). *President George Bush: Election countdown.* Journal Graphics, Denver.

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology, 47,* 1231–1243.

McCauley, C., Durham, M., Copley, J. B., & Johnson, J. P. (1985). Patient's perceptions of treatment for kidney failure: The impact of personal experience on population predictions. *Journal of Experimental Social Psychology, 21,* 138–148.

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

Mullen, B., Atkins, J. L., Champion, D. S., Edwards, C., Hardy, D., Story, J. E., & Vanderklok, M. (1985). The false-consensus effect: A meta-analysis of 115 hypothesis tests. *Journal of Experimental Social Psychology, 21,* 262–283.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90,* 339–363.

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

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46,* 598–609.

Peterson, C. R., & DuCharme, W. M. (1967). A primacy effect in subjective probability revision. *Journal of Experimental Psychology, 73,* 61-65.

Peterson, C. R., Schneider, R. J., & Miller, A. J. (1965). Sample size and the revision of subjective probabilities. *Journal of Experimental Psychology, 69,* 522-527.

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.

Ross, L., Lepper, M. R., Strack, F., & Steinmetz, J. (1977). Social explanation and social expectation: Effects of real and hypothetical explanations on subjective likelihood. *Journal of Personality and Social Psychology, 35,* 817-829.

Rothbart, M. (1981). Memory processes and social beliefs. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 145-181). 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.

Sherman, S. J., Chassin, L., Presson, C. C., & Agostinelli, G. (1984). The role of evaluation and similarity principles in the false consensus effect. *Journal of Personality and Social Psychology, 47,* 1244-1262.

Sherman, S. J., Cialdini, R. B., Schwartzman, D. F., & Reynolds, K. D. (1985). Imagining can heighten or lower the perceived likelihood. *Personality and Social Psychology Bulletin, 11,* 118-127.

Sherman, S. J., Presson, C. C., & Chassin, L. (1984). Mechanisms underlying the false consensus effect: The special role of threats to the self. *Personality and Social Psychology Bulletin, 10,* 127-138.

Tversky, A., & Kahneman, D. (1971). The belief in the law of small numbers. *Psychological Bulletin, 76,* 105-110.

Uleman, J. S., & Bargh, J. A. (1989). *Unintended thought.* New York: Guilford Press.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35,* 151-175.

Zuckerman, M., Mann, R. W., & Bernieri, F. J. (1982). Determinants of consensus estimates: Attribution, salience, and representativeness. *Journal of Personality and Social Psychology, 42,* 839-852.

Received August 26, 1993
Revision received February 3, 1994
Accepted February 11, 1994 ■

New Editors Appointed, 1996–2001

The Publications and Communications Board of the American Psychological Association announces the appointment of two new editors for 6-year terms beginning in 1996. As of January 1, 1995, manuscripts should be directed as follows:

- For *Behavioral Neuroscience,* submit manuscripts to Michela Gallagher, PhD, Department of Psychology, Davie Hall, CB# 3270, University of North Carolina, Chapel Hill, NC 27599.

- For the *Journal of Experimental Psychology: Learning, Memory, and Cognition,* submit manuscripts to James H. Neely, PhD, Editor, Department of Psychology, State University of New York at Albany, 1400 Washington Avenue, Albany, NY 12222.

Manuscript submission patterns make the precise date of completion of 1995 volumes uncertain. The current editors, Larry R. Squire, PhD, and Keith Rayner, PhD, respectively, will receive and consider manuscripts until December 1994. Should either volume be completed before that date, manuscripts will be redirected to the new editors for consideration in 1996 volumes.