

Implicit Gender Stereotyping in Judgments of Fame

Mahzarin R. Banaji
Yale University

Anthony G. Greenwald
University of Washington

Implicit (unconscious) gender stereotyping in fame judgments was tested with an adaptation of a procedure developed by L. L. Jacoby, C. M. Kelley, J. Brown, and J. Jasechko (1989). In Experiments 1–4, participants pronounced 72 names of famous and nonfamous men and women, and 24 or 48 hr later made fame judgments in response to the 72 familiar and 72 unfamiliar famous and nonfamous names. These first experiments, in which signal detection analysis was used to assess implicit stereotypes, demonstrate that the gender bias (greater assignment of fame to male than female names) was located in the use of a lower criterion (β) for judging fame of familiar male than female names. Experiments 3 and 4 also showed that explicit expressions of sexism or stereotypes were uncorrelated with the observed implicit gender bias in fame judgments.

Some of the earliest achievements of information processing methods in social cognition research demonstrated the involvement of memory and other cognitive processes in stereotyping. This research has documented, for example, errors in frequency estimation of behaviors performed by members of varying social groups (Hamilton & Gifford, 1976), superior memory for stereotype-confirming than -disconfirming events (Rothbart, Evans, & Fulero, 1979), and superior memory for information about members of minority than majority groups (see Taylor, 1981). Empirical effects such as these revealed the cognitive basis of stereotyping, and these discoveries changed existing views of the stereotyping process and of the social perceiver. Rather than portraying stereotyping as an intentionally motivated act to discriminate, or as an attribute of a select group of prejudiced individuals, social cognition research has characterized stereotyping as a consequence of the routine operation of cognitive processes.

Staying within this social cognition tradition, the present research investigated stereotyping with an indirect measurement method derived from research on implicit memory, rather than the direct-measurement methods of traditional memory re-

search (viz., recall and recognition). The central interest, however, was not in implicit memory per se, but rather in using a measure of implicit memory to demonstrate the operation of *implicit stereotyping*.¹ This article seeks to establish that methods developed in the study of the unconscious operation of memory can be applied to investigate automatic or unconscious aspects of social cognition. Several recent experiments and discussions have described implicit and automatic cognitive components of stereotypes (Banaji & Greenwald, 1994; Banaji, Hardin, & Rothman, 1993; Bargh, 1994; Devine, 1989; Dovidio, Evans, & Tyler, 1986; Fiske, 1989; Geis, 1993; Gilbert & Hixon, 1991; Perdue & Gurtman, 1990; Pratto & Bargh, 1991). The present experiments extend this work in two ways. First, they demonstrate the use of a specific method to study implicit memory (the false-fame procedure) that has potentially broad further application in the study of stereotypes. Second, they introduce the use of signal detection theory and methods (Green & Swets, 1966) to distinguish two cognitive components of stereotyping.

From Implicit Memory to Implicit Stereotyping

A distinguishing feature of implicit memory phenomena is that the trace of a past experience influences later performance, even though the earlier experience is not remembered in the usual sense—that is, it is unavailable to self-report or introspection (cf. Graf & Schacter, 1985; Jacoby & Dallas, 1981; Jacoby,

Mahzarin R. Banaji, Department of Psychology, Yale University; Anthony G. Greenwald, Department of Psychology, University of Washington.

This research was supported by National Science Foundation Grants DBC 9120987 and DBC 9205890 and National Institute of Mental Health Grant MH 41328. We thank R. Bhaskar, Curtis Hardin, John Jost, Myron Rothbart, and Alexander Rothman for comments; Chris Crandall, Robert Crowder, and Judith Primavera for providing opportunities for classroom testing; and Coomi R. Banaji and Wendi Walsh for assistance with the manuscript.

Correspondence concerning this article should be addressed to Mahzarin R. Banaji, Department of Psychology, Yale University, P.O. Box 208205, New Haven, Connecticut 06520-8205. Electronic mail may be sent via Internet to mbanaji@yalevm.cis.yale.edu.

¹ When we began this research, we intended to study attitudes that operated outside of conscious control. The result of the first experiment in this series was reported by Greenwald (1990) and Banaji and Greenwald (1991) as demonstrating the implicit operation of *attitudes*. After further experiments and consideration of the relation of those experiments' findings in relation to other findings, it became clear that the research was better described as involving the unconscious operation of *stereotypes*.

Lindsay, & Toth, 1992; Jacoby & Witherspoon, 1982; Roediger, Weldon, & Challis, 1989; Schacter, 1987). As an illustration of implicit memory, consider a result that is readily obtained with the task of identifying words that are presented as brief tachistoscopic flashes. Successful identifications of words occur more often for words to which participants were casually exposed earlier in the experiment than for new words that were not previously presented. This effect of prior exposure occurs despite participants' poor performance on tasks of recalling or recognizing words from the earlier list. The tachistoscopic identification task provides an indirect measure of the effect of the prior experience. That is, even though the participant is not instructed to retrieve the earlier presented material and is presumably not trying to do so (and may well be incapable of such retrieval), the participant's responses indicate a residual effect. (For further reviews of indirect measurement in memory research, see Richardson-Klavehn & Bjork, 1988; Roediger, 1990; Roediger & Blaxton, 1987; Tulving & Schacter, 1990.)

Extending beyond the memory literature, Greenwald and Banaji (1995) identified *implicit social cognition* as occurring when past experiences affect social judgment or behavior, but the nature of this influence is introspectively unidentified (or inaccurately identified) by the actor. They offered examples of several implicit social cognition phenomena, focusing on the research on attitudes, self-esteem, and stereotyping. In each of these domains, Greenwald & Banaji (1995) (a) showed that social behavior operates in an implicit or unconscious fashion and (b) called for greater use of indirect measures of social cognition. Applying the definition of implicit social cognition to the domain of stereotyping, they defined implicit stereotypes as *introspectively unidentified (or inaccurately identified) traces of past experience that mediate attributions of qualities to members of a social category*. Such a definition has new implications for both theory and research. For theory, it provides explicit recognition of the unconscious manner in which stereotypic expressions may be produced and influence judgment and action, and the possible dissociation between conscious and unconscious components of stereotyping. For method, the implications are substantial. Such a definition highlights the previously ignored role of unconscious processes in dominant measures such as adjective checklists and rating scales, and encourages the use of new, indirect measures in investigations of stereotyping.

Obvious attributes that distinguish members of social categories are race and ethnicity, gender, age, class, and correlated attributes of skin color, physical features, dress, accent, physical abilities, and personality attributes. A goal of research on implicit stereotyping is to identify the unconscious use of such attributes in social judgment and the consequences of such judgments for target and perceiver.

Gender is known to be associated with strong and widely shared stereotypes (Basow, 1986; Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972; Eagly & Mladinic, 1989; Ruble & Ruble, 1982; Williams & Best, 1990). A well-known gender difference involves achievement, with men more than women being associated with achievement in many fields, such as politics, sports, and science. One source of data revealing the gender-specific association of achievement is from archives in-

dicative of fame: Men constitute 93% of the entries in *Who's Who in America* (1988), which purports to identify outstanding achievers in such arenas as business, government, arts, science, law, and education. Additionally, as a barometer of popular culture, a board game called Claim to Fame, which tests the ability to recognize famous people, shows the imbalance of having a 3.85:1 male:female ratio among its 912 famous names.²

The discrepancy in male and female fame in archival and cultural indicators is matched by the cognitive availability of male names compared with female names. We asked Yale undergraduates to generate the first instance that came to mind in response to various categories (e.g., a book, a piece of furniture, an article of clothing) including "a famous person" and found that, of the 87 participants, 73 (84%) generated a male name, and only 14 (16%) generated a female name (a 5.21:1 ratio). Thus, discrepancies in male-female fame, which are well established in archival data, are similarly indicated by a straightforward measure of cognitive availability.

A Procedure to Test Implicit Gender Stereotyping

In the present experiments we used an implicit memory procedure developed by Jacoby, Kelley, Brown, and Jasechko (1989) to observe implicit effects of prior exposure to names on subsequent fame judgments. In one experiment, Jacoby et al. (1989) exposed participants to names of famous and nonfamous people by asking them to pronounce each name. Then, either immediately or after a 24-hr delay, participants were presented with previously seen (old) and previously unseen (new) famous and nonfamous names and were asked to make the following judgment for each name: Is this person famous? Jacoby et al. (1989) hypothesized that, in the delayed judgment condition, episodic memory for the familiar names would dissipate, but perceptual fluency for the name would persist. This dissociation of (explicit) episodic memory from (implicit) perceptual fluency for the names was predicted to lead to a higher probability of incorrect judgments of familiar nonfamous names as famous, compared with otherwise equivalent unfamiliar nonfamous names. In other words, the familiarity evoked by a name due to seeing it recently was confused with the familiarity associated with a famous name. Consistent with this expectation, Jacoby et al. (1989) obtained a reliably higher rate of false alarms (i.e., false fame) for familiar nonfamous than unfamiliar nonfamous names in their delayed-judgment condition.

In the present research we adapted the false-fame procedure to examine a hypothesized gender bias in fame judgments. Names communicate social information, such as race and ethnicity, social class, age, and competence (Allport, 1954; Goldberg, 1968; Kasof, 1993), but especially gender. In the present experiments, names were very useful stimuli because they allowed the category of gender to be presented without making it

² The purpose of reviewing these statistics is merely to point out that differences in male and female fame exist, while bypassing discussion of why they may exist. As compilers of women's biographical dictionaries point out, the difference in fame is likely to be in part a reflection of historical and present-day reality and in part a reflection of a male bias in the writing of history and the reporting of current events (McHenry, 1980).

obvious that an effect of gender was being investigated. Participants were exposed to names of men and women by initially having them make a pronounceability judgment on each name. Later, participants were shown both familiar (previously seen) and unfamiliar (new) famous and nonfamous names and asked to give a dichotomous (famous–nonfamous) judgment on each name.

Jacoby et al. (1989) regarded the perceptual fluency with the name to be critical in obtaining the fame bias. Prior exposure, according to their theory, causes the fluency that serves as the basis for the (incorrect) judgment of fame. We assumed that, after initial exposure, participants would acquire equal perceptual fluency with male and female names. Therefore, if, at the time of judgment, perceptual fluency alone determines judgment, equal false fame effects should emerge for male and female names. If, however, the judgment of fame is sensitive to prior knowledge of the gender–fame association—an aspect of the gender stereotype³—familiarized male names should be more likely to be judged as famous than familiarized female names. As a replication of previous research, the false-fame bias should be obtained in the form of greater false alarms for old than new names. The gender bias in false fame may be obtained in the form of greater false alarms for familiarized male than female names. However, as illustrated in the next section, the gender bias in false fame will be examined more thoroughly by computing the two statistics associated with signal detection theory. These statistics, we argue, map onto two theoretically distinct processes that underlie stereotyping.

Components of Stereotypes: Theory

Stereotypes can be described as having the form *Xs believe that Ys are especially Z* (e.g., Americans believe that elderly people are especially poor in memory). Regardless of the extent to which a genuine association of *Y* with *Z* exists, the application of a stereotype has been assumed to produce a distortion by ignoring characteristics of the individual in favor of a belief about the group as a whole. For example, the stereotype is assumed to reduce ability to perceive distinctions in memory ability among individual elderly persons compared with typical ability to notice such distinctions among younger people. This hypothesized property of stereotypes can be identified as *reduced sensitivity*. Reduced sensitivity is widely considered to be apparent in comparison of ingroup versus outgroup perceptions, with perceivers being less sensitive to differences among outgroup than ingroup members—the *outgroup homogeneity effect* (Hewstone, Islam, & Judd, 1993; Linville, Salovey, & Fischer, 1986; Ostrom & Sedikides, 1992; Park & Rothbart, 1982).

A possible alternative view is that stereotypes operate without altering sensitivity to individual characteristics. Rather, they may operate by consistently shifting judgments—that is, displacing them away from presumably veridical values in the stereotypic group-characteristic direction. For example, the stereotyped association of old age with poor memory might result in a consistent lowering of the estimate of a person's memory ability based on knowledge that the person is old. This second theorized property of stereotypes can be identified as *bias*.

The distinction between reduced sensitivity and bias as operational properties of stereotypes is important in attempting to

undo the perceptual distortions of stereotypes. For example, social workers may be aware that their stereotypes can lead them to misjudge the memory ability of elderly clients. In attempting to avoid or compensate for such expected errors, it would be useful for them to know either that their initial estimates are generally likely to be too low because of a bias, or that their estimates are likely to underestimate differences among clients because of reduced sensitivity. Errors due to bias may be easier to correct than are errors due to reduced sensitivity.

Components of Stereotypes: Method

The distinction between reduced sensitivity and bias as properties of stereotypes corresponds to a distinction that is fundamental to the method of signal detection analysis in perception research (Green & Swets, 1966). Signal detection analysis is used to decompose hit and false alarm rates for two categories of stimuli (signal and noise, respectively) into measures of the perceiver's sensitivity to signal presence and the perceiver's threshold or criterion for assigning the judgment. These measures are identified respectively as d' (d-prime, the measure of sensitivity) and β (beta, the measure of criterion). Variations in the perceiver's criterion are commonly referred to also as variations in bias (that is, bias to judge the stimulus as signal rather than noise).

Consider the application of signal detection methods to the false-fame experiment. The false-fame experiment uses two categories of stimuli (famous and nonfamous names), which correspond to the standard signal and noise categories, respectively. For each participant, the hit rate is the proportion of famous names that he or she correctly identifies as such, whereas the false alarm rate is the proportion of nonfamous names that the participant mistakenly identifies as famous. Signal detection analysis permits converting these two rates into a d' measure that indicates the participant's sensitivity to actual fame, and a β measure that indicates the participant's criterion for (or bias in) assigning the fame judgment. That is, d' indicates a participant's sensitivity to fame of names—the ability to accurately report the difference between famous and nonfamous names. On the other hand, β is a measure of the strictness of the decision rule that the participant applies in judging each name as famous or nonfamous.

The main hypothesized result of greater false fame for previously seen male versus female names is assumed to be produced by the co-occurrence of two conditions, both of which are seen as necessary to produce the effect: first, some degree of fluency for the perceptual features of the names, and second,

³ Stereotypes have been defined in various ways, with earlier definitions focusing on the incorrect or exaggerated nature of beliefs about social groups (Allport, 1954; Katz & Braly, 1935). Consistent with other definitions (e.g., Aronson, 1988; Secord, 1959), we regard stereotyping to have occurred whenever knowledge (accurate or inaccurate) about a social category is used in judging an individual member of the category. Thus, in the present research, gender stereotyping will be said to be involved when the attribute of fame is conferred more readily on male names than on equally familiar female names, even though the differential judgment may be justified by the actual distribution of fame in populations of men and women.

the availability of knowledge about the association between gender and fame in the judging of each name. In the present variation of the false-fame experiment, male versus female gender of names is thus included as an additional independent variable, meaning that d' and β can be determined separately for male and female name stimuli. Use of signal detection methods, therefore, permits description of the stereotypic association of gender with achievement in terms of differences in sensitivity, bias, or both, between male and female names. The simplest finding would take the form of greater readiness to assign the judgment of fame to male than to female names. This finding would indicate differential criteria for judging fame for male and female names, with the prediction of a stricter criterion applied to female than male names.

The signal detection analysis also permits the conclusion that the perceptual difference consists exclusively of a sensitivity difference. Such a finding would indicate a difference in discriminability of famous from nonfamous names for male versus female names. If such a sensitivity difference is obtained, the direction of difference might reveal lower sensitivity for male names—that is, greater fame accorded to male names may stem from reduced ability to separate famous male from nonfamous male names. Of course, these experiments may also show that both sensitivity and bias effects underlie gender stereotyping. The power of signal detection analyses comes from the opportunity to separate bias from sensitivity effects and to test the contribution of both to producing the gender bias in false fame. In addition, these analyses will inform about the nature of the false-fame bias itself. Jacoby et al. (1989) provided only false alarm data, and hence it is not known whether the false-fame bias is rooted in sensitivity or bias.

Added to these possibilities is the use of both male and female participants in the present research, which allows a measure of differences in implicit stereotypes held by men and women. The research on outgroup homogeneity should predict that men should show lower sensitivity for female names, and women should show lower sensitivity for male names. However, if these outgroup homogeneity effects are restricted to explicit stereotyping, such gender differences should not be obtained on implicit stereotyping measures.

Experiment 1: A Gender Bias in Judgments of False Fame

Experiment 1 provided an initial test of a gender bias in false fame, using signal detection methods. The procedure used by Jacoby et al. (1989) was revised to include an equal number of male and female famous and nonfamous names, and exposure to the names and the final judgment task were implemented with a paper-and-pencil version of Jacoby et al.'s computerized presentation. In the second session, 24 hr later, participants were provided with twice as many names, including all of the first-session famous and nonfamous names (familiar names) and an equal number of unfamiliar names of the same types and were asked to judge each for fame.

Method

Participants

Thirty high school juniors and seniors enrolled in a summer section of Introductory Psychology at Yale University participated in partial

fulfillment of a course requirement. Participants were tested in the classroom in two sessions, 24 hr apart.

Materials

Famous names. In their research, Jacoby et al. (1989) used only a small proportion of female names (approximately 20%). For the present research, we developed lists of 12 famous men and 12 famous women in each of six categories: actors, writers, musicians/singers, scholars, politicians, and athletes. All names were selected such that (a) each name was judged to be famous enough that most college student participants could recognize it as famous, but not so famous that most participants could specify the achievement that led to fame, (b) highly distinctive last names were excluded (for example, Goolagong), and (c) names that contained a middle initial or name were also excluded (for example, Sandra Day O'Connor, Billie Jean King). Examples of famous female and male names that were used are: Jane Wyman, Gladys Knight, Doris Lessing, Margaret Mead, Bella Abzug, Peggy Fleming, Rod Steiger, Miles Davis, Thornton Wilder, Arnold Toynbee, Hubert Humphrey, and Mark Spitz.

Nonfamous names. After we selected the 144 famous names, we derived 288 nonfamous names from them. We constructed a matched nonfamous male or female name for each male or female famous name by retaining the famous last name and adding an appropriate first name. We used the following criteria to construct nonfamous first names from famous ones: (a) nonfamous male and female names were matched for number of letters, (b) the first letter of a nonfamous first name always differed from the first letter of the famous first name, (c) if the famous first name was a nickname, the nonfamous version was also a nickname, and (d) gender-ambiguous first names (e.g., Robin, Leslie) were not used. Examples of nonfamous female and male names that were used are: Anne Wyman, Lauren Knight, Susan Lessing, Bret Curtin, David Wharton, Danny Holiday, Joe Steiger, Jacob Davis, John Spitz, Sandy Riggs, Hannah Jackson, and Ella Kemp.

Name lists. We constructed two sets of 216 names (Set A and Set B); each set consisted of 72 famous names, 72 nonfamous male derivatives, and 72 nonfamous female derivatives. The 72 famous names consisted of 12 names (6 male, 6 female) in each of six categories. From each of Sets A and B, we created four lists, (A1–A4; B1–B4), each of which contained 36 famous and 36 nonfamous names (18 male and 18 female in each). Thus, in each set, each of the four lists contained the same 72 different last names. A subset of 36 famous names appeared identically in two of each set of four lists; these two lists differed in the gender of nonfamous first names that accompanied the remaining 36 (famous) last names. Each participant received a 72-name list from one set (e.g., A1) in Session 1, and that list plus the corresponding list from the other set (e.g., A1 + B1) in Session 2. This procedure always ensured that familiar and unfamiliar names never contained the same last names.

Overall, in Session 1, each participant received 72 names: 18 famous female, 18 famous male, 18 nonfamous female, and 18 nonfamous male. In Session 2, each participant received 144 names: 72 familiar and 72 unfamiliar ones. The 72 unfamiliar names were identical to the 72 familiar ones in fame and gender composition. For the eight different booklets of 72 names used in Session 1, and the four different booklets of 144 names used in Session 2, the names appeared in random order, with two constraints. No more than 4 male or female names, and no more than 4 famous or nonfamous names, appeared in sequence.

Design

The design for names presented in Session 2 was a 2 (name gender: male or female) \times 2 (name exposure: familiar or unfamiliar) \times 2 (participant sex: male or female) factorial, with only participant sex being

a between-subjects variable. The independent variable of name fame (famous or nonfamous) is not specified as a design variable, rather being incorporated into the signal detection dependent variables, d' and β .

Procedure

In Session 1, each participant received a list of 72 names preceded by 2 buffer names (1 famous, 1 nonfamous). Participants were informed that the purpose of the study was to obtain ratings of the ease of pronounceability of common names. Twenty-six names appeared on the first page, and 24 names appeared on each of two additional pages. Each name was accompanied by a 5-point scale (1 = *very easy to pronounce* through 5 = *very hard to pronounce*). Participants were asked to read each name silently and to rate the ease of pronouncing the name "for a person who speaks little English." The only purpose of the pronounceability judgment task was to provide familiarity with the names.

In Session 2, participants judged whether each name was famous or not by circling either the word "Famous" or "Nonfamous" next to each. Finally, participants provided information about the number of years they had resided in the United States and their native language. These data were collected to provide a basis for eliminating any participants who were newly arrived in the United States and may therefore have been unfamiliar with the experiment's sample of famous names. However, no participants needed to be dropped on the basis of these responses. Participants were debriefed about the purpose of the study, after which the experimenter engaged them in conversation about various aspects of the two tasks. No participant reported ideas resembling the gender stereotype hypothesis of the experiment and, after learning of it, several communicated doubt that their data would reveal any gender bias in fame judgment.

Results

Data from each participant's judgments of the 18 famous and 18 nonfamous names for each of the four within-subject conditions (familiar male, familiar female, unfamiliar male, and unfamiliar female names) were reduced to a hit rate (proportion of famous names correctly judged as famous) and a false alarm rate (proportion of nonfamous names mistakenly judged as famous). Two participants were dropped from all analyses, 1 because of an unusually high number of false alarms (13 out of 18 possible) in the familiar male names condition and another for an unusually low hit rate across all name conditions (only 5 out of 72 correct). (Inclusion of these 2 participants in analyses, however, did not alter the pattern of statistically significant and nonsignificant findings from those reported below.) All analyses were first performed using participant sex as a variable, but because no effect of participant sex was obtained in any analysis (all F s < 1), that variable was dropped from the reported analyses.

As can be seen in Figure 1, hit rates were higher for familiar names than for unfamiliar names, $F(1, 27) = 8.08, p = .008$, and hit rates were higher for male names than for female names, $F(1, 27) = 22.64, p < .0005$, with the interaction effect of name familiarity (familiar or unfamiliar) and name gender (female or male) being nonsignificant, $F(1, 27) = 0.26$. False alarm rates were higher for familiar names than for unfamiliar names, $F(1, 27) = 7.16, p = .012$, replicating the result obtained by Jacoby et al. (1989). Also, the false alarm rate for male names was higher than that for female names, but only marginally so, $F(1, 27) = 4.11, p = .053$, and the interaction effect of name famil-

ilarity and name gender on false alarm rates—although in the predicted direction—did not reach statistical significance, $F(1, 27) = 3.14, p = .088$.⁴

We used signal detection methods (Green & Swets, 1966) to decompose the hit and false alarm rates into estimates of each participant's sensitivity (d') to the famous–nonfamous distinction and criterion (β) for using stimulus information on the fame dimension to arrive at a judgment of fame. The main hypothesis tests used these measures, replacing β with its logarithm because of that transformation's better approximation to a normal distribution. Signal detection analyses require that hit and false alarm rates be neither zero nor 100%. An adjustment was made for empty cells (e.g., 0% false alarms or 100% hits in one or more conditions for a participant).⁵

Mean values of d' and $\log \beta$ for the four conditions are provided in Figures 2 and 3, respectively. The results for d' indicate that participants did not differ in sensitivity to the fame variation for familiar versus unfamiliar names, $F(1, 27) = 2.12, ns$, or for male versus female names, $F(1, 27) = 0.43, ns$. The results for $\log \beta$ showed that participants more readily judged familiar than unfamiliar names to be famous, $F(1, 27) = 16.01, p < .0005$. This result suggests that the effect obtained by Jacoby et al. (1989) was likely due to a lower criterion for judging fame in familiar than in unfamiliar names. More central to the present aim of investigating stereotypes, analyses of $\log \beta$ showed that participants used a more lenient criterion for judging fame in male than female names, $F(1, 27) = 10.81, p = .003$.

In examining interaction effects, statistically significant effects were found for both d' and $\log \beta$. For d' the interaction effect appeared as greatest sensitivity to unfamiliar male names than to the other categories, $F(1, 27) = 5.08, p = .032$. For $\log \beta$, the most apparent feature of the interaction effect was an especially low criterion for familiarized male names, $F(1, 27) = 4.73, p = .039$. Because this interaction is of greatest theoretical interest, we report the most important simple main effects. The benefit of false fame accrued more to familiar male than familiar female names, $t(27) = 3.46, p = .002$, whereas such a difference for male and female names was not obtained on unfamiliarized names, $t(27) = 1.40, p = .17$. All remaining statistical

⁴ The criterion of statistical significance throughout this report is $\alpha = .05$, two-tailed.

⁵ Averaged over the four experiments, the signal detection analyses reported in this article required dealing with 88% of participants for whom we had one or more empty cells in the 2⁴ classification of responses as signal (famous name) versus noise (nonfamous name), male versus female, familiar versus unfamiliar, and judgment of name (famous vs. nonfamous). The presence of a single empty cell in a participant's data yielded a computationally obstructive infinity when the observed proportion of 0 was interpreted as an area to be converted to a standard normal deviate for computation of d' or $\log \beta$. In the course of the present research we developed a method to deal with the computational problem caused by empty cells, and we compared this procedure with other existing fixup methods for dealing with empty cells. Although this method is likely to be useful in other situations in which empty cells disrupt proportion- or ratio-based computations, no attempt was made here to establish its value in a general fashion. More information about the details of the particular method developed for statistical adjustment for empty cells in analyzing proportions may be obtained by writing to the authors.

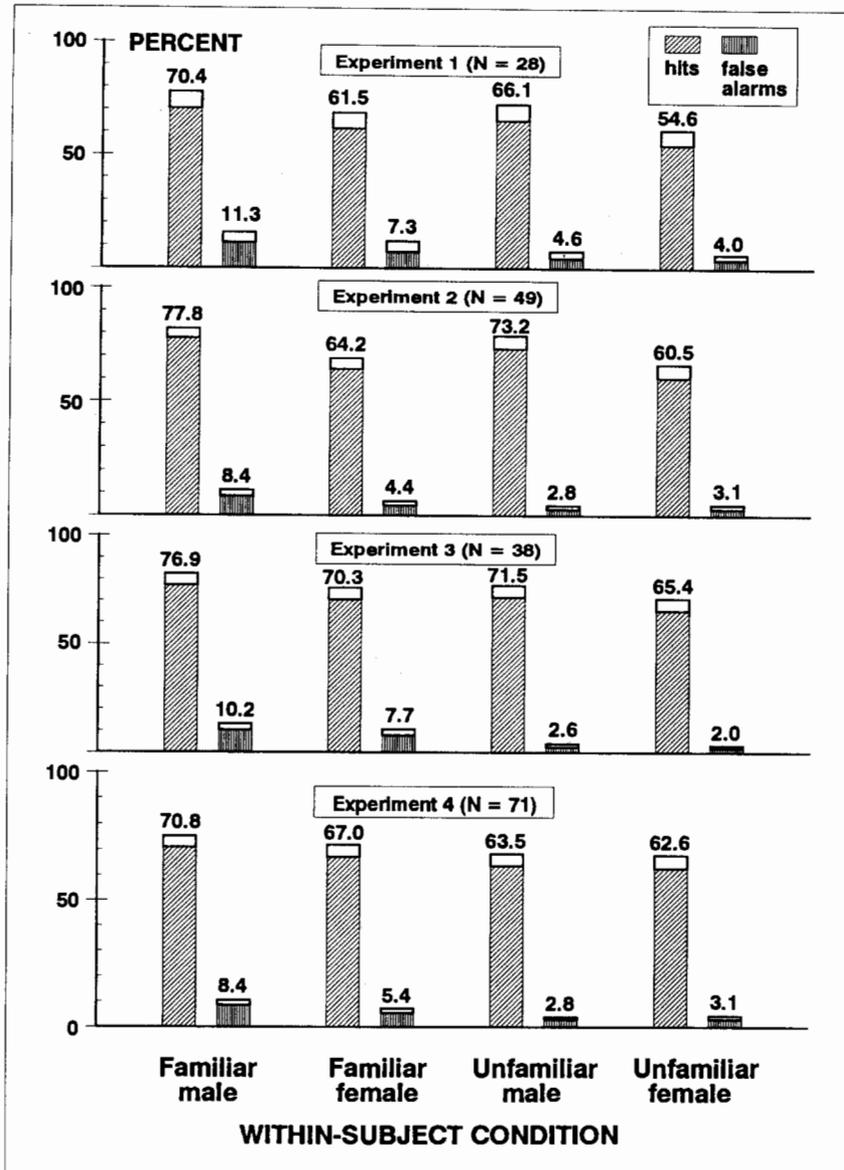


Figure 1. Mean hit rates and false alarm rates as a function of within-subject conditions of name familiarization and gender. The unfilled extensions of bars show the upper half of the 95% confidence interval for each mean.

decompositions considering patterns of statistical significance in two-treatment contrasts among the four means is not given here. Rather, this more detailed analysis is deferred to the General Discussion section, where we present a combined analysis of the four experiments. This strategy, which we also used for Experiments 2, 3, and 4, serves to avoid populating each experiment's *Results* section with multiple significance tests that may be rendered obsolete by the more powerful summary analysis.

Discussion

The results of Experiment 1 indicated that gender stereotypes operated with greatest force for nonfamous names that were

given a boost in familiarity by their presentation in the first session. When encountered in the second session, the familiarity of these nonfamous names was more likely to be interpreted as arising from fame when the name was male than when it was female. In Experiment 1, using a paper-and-pencil version of Jacoby et al.'s (1989) procedure, we replicated their finding that familiar nonfamous names are indeed more likely to be incorrectly judged as famous than comparable, new nonfamous names. The signal detection analysis further indicated that participants did not differ in their sensitivity (d') to fame for male and female names but did discriminate between male and female names by having a lower threshold or criterion (β) for assigning the judgment of fame to male names.

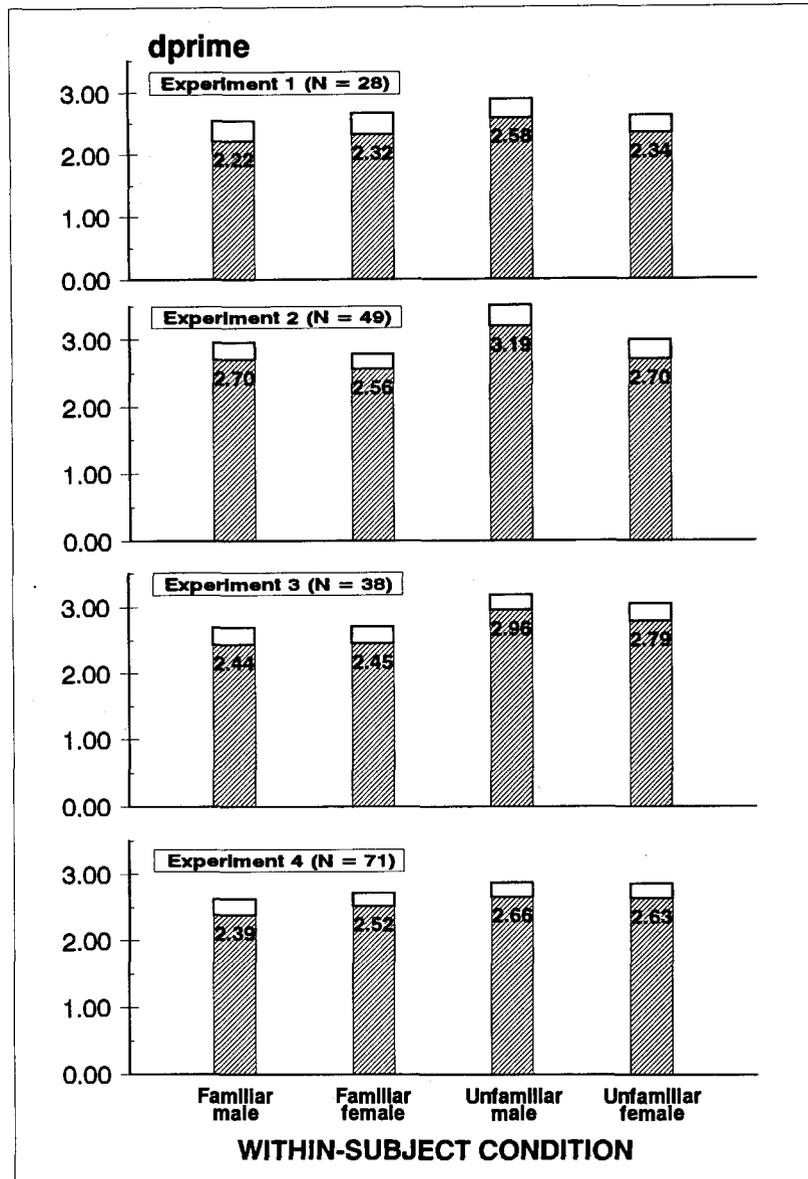


Figure 2. Mean d' scores as a function of within-subject conditions of name familiarization and gender, showing upper halves of the 95% confidence interval for each mean.

This finding of a gender bias in the threshold for fame judgments qualifies as an implicit stereotyping effect. Although the experiment provided no direct evidence that the observed gender bias in fame judgments occurred without participants' awareness, debriefing procedures in this and other experiments indicated that participants almost never showed awareness that (a) gender of the name was of interest to the experimenter, and (b) gender of name had influenced their judgment of fame.

Experiment 2: A Replication With Variation in Stimuli and Delay

In Experiment 1, nonfamous names were created by assigning new male and female first names to famous male and

female last names (e.g., Merrill Wilder and Roberta Wilder from Thornton Wilder). It was possible that the gender origin of nonfamous names influenced their likelihood of being (falsely) judged famous. To test this, each false alarm for each participant was placed in one of eight categories to test whether false alarms occurred more often on nonfamous names that had been derived from a famous name of the same gender. We constructed the eight categories by identifying each nonfamous name as familiar (previously seen) or unfamiliar (unseen), male or female first name, and male or female root name. The number of false alarms in each category was divided by the participant's total number of false alarms to derive a proportion of false alarms for each of these eight categories.

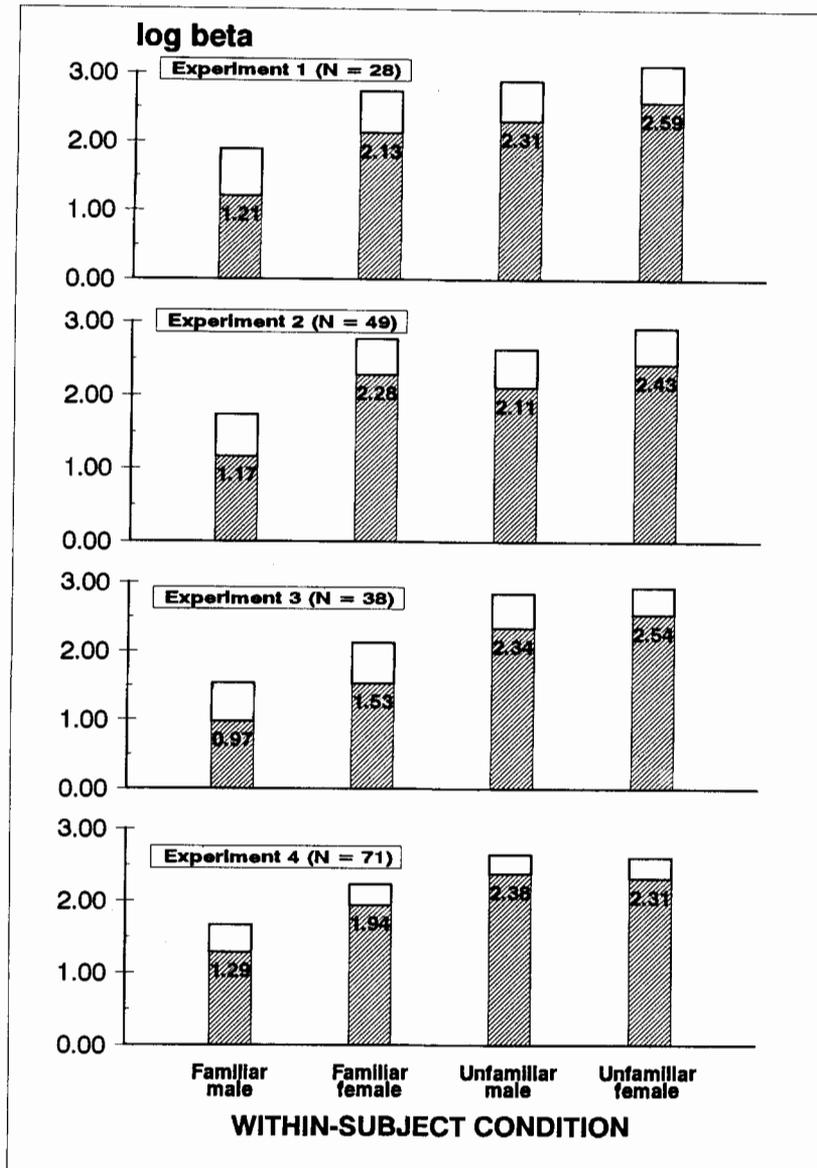


Figure 3. Mean $\log \beta$ scores as a function of within-subject conditions of name familiarization and gender, showing upper halves of the 95% confidence interval for each mean.

The main result apparent in this analysis was that more false alarms occurred on names derived from male names (64%) than on ones derived from female names (36%). This observation led to a revised procedure for obtaining nonfamous names in Experiment 2, the main purpose of which was to provide a replication that could bolster the finding of implicit gender stereotyping from Experiment 1. The famous names used in Experiment 2 came from three of the six categories used in Experiment 1: writers, actors, and musicians. The other three categories (scholars, athletes, and politicians) were dropped because of relative difficulty in identifying women who matched men in actual fame. The only other substantial change was the use of a 48-hr (rather than 24-hr) delay between sessions, neces-

sitated by the meeting schedule of the class that provided participants. Because of previous findings that perceptual fluency effects survive extended periods between exposure and test (Kolers, 1976), this increased delay of Session 2 was not expected to alter findings.

Method

Participants and Design

Participants were tested in their classroom during both sessions. Fifty-seven students in an introductory psychology class participated in Session 1. Of these participants, data were used for only 49 (23 men and 26 women) of the 50 who were also present during Session 2, 48 hr later.

Data for 1 male subject were discarded because of an unusually high rate of false alarms on nonfamous names (23%, approximately equal for male and female names, compared with a mean of 5.5% for other participants). The design was a 2 (name gender: male or female) \times 2 (name exposure: familiar or unfamiliar) \times 2 (participant sex: male or female) factorial, with only participant sex as a between-subjects variable.

Materials and Procedure

From the famous names used in Experiment 1, 12 famous names in each of these three categories (writers, musicians, and actors) were used. Twelve additional names were added to each category to create a total of 24 famous names (12 female and 12 male) in each. The resulting set of 72 famous names was divided into two lists of 36, each containing 12 names (6 male and 6 female) in each of the three fame categories.

Nonfamous last names were selected from the New Haven, Connecticut phone book, each matching one of the 72 famous names in length and ethnicity. To each nonfamous last name was associated both a female and male first name, each coordinated with the famous name in the same fashion as previously described for Experiment 1. The full set of 144 nonfamous names thus included 72 different last names, each associated with both a male and a female first name (e.g., Jane Smith, John Smith, Judy Martin, Andy Martin).

We divided the 144 nonfamous names into four lists of 36 names each (18 male, 18 female). Lists A1 and A2 shared the same last names, with the only difference that last names in List A1 that were attached to a female first name were attached to a male first name in List A2, and vice versa. The same relationship existed between lists B1 and B2. Participants who got an A list in Session 1 always received a B list as the new (unfamiliar) names in Session 2 (and vice versa), assuring no 2 nonfamous names shared the same last name. As in Experiment 1, each participant received 72 names in Session 1 (18 each of famous female, famous male, nonfamous female, and nonfamous male), and 144 names in Session 2 (the 72 familiar Session 1 names, along with 72 unfamiliar ones similarly composed).

Results and Discussion

As in Experiment 1, data from the four within-subjects conditions (familiar male, familiar female, unfamiliar male, unfamiliar female) were reduced to hit and false alarm rates. As Figure 1 shows, hit rates were higher for male names than for female names, $F(1, 48) = 85.10, p < .0005$, and were higher for familiar names than for unfamiliar names, $F(1, 48) = 5.22, p = .027$, with the interaction effect being nonsignificant, $F(1, 48) = 0.09$. False alarm rates were higher for familiar names than for unfamiliar names, $F(1, 48) = 15.51, p < .0005$, and for male names than for female names, $F(1, 48) = 9.81, p = .003$. A significant interaction of name familiarity and name gender took the form of an especially high false alarm rate for familiarized male names, $F(1, 48) = 10.55, p = .002$.

We performed signal detection analyses again, as for Experiment 1. As before, there were no significant effects of participant sex on data for either d' or $\log \beta$, and consequently that between-subjects variable was dropped from the analyses presented here. Mean values of d' and $\log \beta$ are presented in Figures 2 and 3, respectively. The results for d' indicate that participants were more sensitive to the fame variation for familiar than for unfamiliar names, $F(1, 48) = 11.15, p = .002$, and more sensitive to the fame variation for male than female names, $F(1, 48) = 17.01, p < .0005$. The interaction effect of name familiarity

and name gender on d' was not significant, $F(1, 48) = 3.71, p = .06$. The results for $\log \beta$ showed that participants more readily assigned the judgment of fame to familiar names than to unfamiliar names, $F(1, 48) = 13.36, p = .001$, again replicating the effect observed by Jacoby et al. (1989). Relevant to the implicit stereotyping hypothesis, participants more easily attributed fame to male than to female names, as evidenced by the lower $\log \beta$ for male than female names, $F(1, 48) = 19.33, p < .0005$. Also, the interaction effect of name familiarity and name gender was significant, $F(1, 48) = 6.33, p = .015$, again taking the form of an especially low criterion for the condition of familiarized male names. The crucial comparison on $\log \beta$ again indicated that familiar male names were more likely to be judged as famous than familiarized female names $t(48) = 4.23, p = .0001$, but the same difference for unfamiliarized names was not statistically significant $t(48) = 1.75, p = .08$.

The main finding of Experiment 2 is its replication of Experiment 1's evidence for gender bias in fame judgments in the form of a lower criterion for judging a male name as famous. The experiment also found a difference in d' for male versus female names, which was not found in Experiment 1. This result could be an artifact of the materials used for the research—in particular, male and female names that are objectively unequal in fame—and is monitored in the next two experiments, in which we used the same stimulus names.

Experiment 3: Exploring the Relationship Between Explicit Beliefs and Implicit Stereotyping

In addition to providing a replication test of the gender bias in fame judgment of Experiments 1 and 2, in Experiment 3 we examined whether the individual differences in implicit gender stereotyping, as indicated by errors in fame judgments, are reflected also in participants' explicit beliefs about gender equality. The comparison of implicit with explicit evidence for stereotyping can indicate whether these are dissociated from one another, comparable to the dissociations sometimes observed between implicit and explicit memory (see Roediger et al., 1989). If a dissociation between explicit and implicit measures is found, it will suggest that the two types of measures may not measure the same underlying construct, and this result, in conjunction with the heavy reliance on explicit measurement techniques in social cognition (see Greenwald & Banaji, 1995), will question the wisdom of depending solely on explicit measures.

Previous research suggests the possibility that explicit (conscious) and implicit (unconscious) indicators of stereotyping may be dissociated. Several investigators have observed that people who can be distinguished as prejudiced versus nonprejudiced on direct measures may not be distinguishable on indirect measures (Crosby, Bromley, & Saxe, 1980; Devine, 1989; Gaertner, 1973). These earlier demonstrations have been based on studies of special populations of participants who represent extremes of explicit measures of race prejudice or political ideology. In the present research, no attempt was made to select participants who fall at extremes of direct measures.

Method

Participants and Design

38 upperclass students (25 women, 13 men) enrolled in a course on learning and memory participated in the experiment, in two sessions 48

hr apart. For both sessions, participants were tested in their classroom. The design was a 2 (name gender: male or female) \times 2 (name exposure: familiar or unfamiliar) \times 2 (participant sex: male or female) factorial, with all but participant sex as within-subjects variables.

Materials and Procedure

The materials (name lists) and procedure were identical to those of Experiment 2, with the addition of the explicit measure of sexism following Session 2's fame judgment task. In measuring explicitly held beliefs about gender equality, two scales have seen frequent use in recent research: the Feminism Scale (FEM; Smith, Ferree, & Miller, 1975) and the Attitudes Toward Women Scale (ATW; Spence & Helmreich, 1972). Unfortunately, both the FEM and ATW scales contain items that appear to be outdated for use with samples of contemporary college students (for example, a FEM item reads, "The 'clinging vine' wife is justified provided she clings sweetly enough to please her husband"). Pilot data indicated that some participants simply refused to answer several items, and most students were clustered at the feminist or belief-in-equality ends of the scales. Instead of using these existing measures, in Experiment 3 we used a set of face-valid items that we had selected on the basis of both their meaningfulness to the present participant population and the variability of responses they elicited.

Eleven participants generated several statements representing a range of opinions about gender equality. From the resulting pool of more than 30 statements, we selected 8 that we considered to reflect gender issues of interest to students. Of these, a "strongly agree" response to 4 of them (Numbers 1, 3, 5, and 7) indicated beliefs in gender equality (feminism), and a "strongly agree" response to the remaining 4 (Numbers 2, 4, 6, and 8) indicated endorsement of gender inequality (sexism).⁶ Participants made their judgment for each item on a 5-point Likert scale with options labeled Strongly Agree, Agree, Neutral, Disagree, and Strongly Disagree. Items were scored in the direction of sexism, such that the greater the sexist endorsement, the higher the score. The overall sexism score was computed as the mean score across these 8 items. An additional, separate dichotomous score was obtained from the ninth item, which asked whether or not the participant accepted a self-description of *feminist*.

Results

As in Experiments 1 and 2, we classified data from judgments of fame into four within-subjects categories (familiar male, familiar female, unfamiliar male, unfamiliar female) and reduced them to hit and false alarm rates. As Figure 1 shows, hit rates were higher for male names than for female names, $F(1, 37) = 16.17, p < .0005$; and for familiar names than for unfamiliar names, $F(1, 37) = 6.80, p = .013$; with the interaction effect being nonsignificant, $F(1, 37) = 0.02$. The false alarm rate was higher for familiar names than for unfamiliar names, $F(1, 37) = 39.32, p < .0005$; and for male names than for female names, $F(1, 37) = 5.15, p = .029$; with the interaction effect of name familiarity and gender being nonsignificant, $F(1, 37) = 1.04$.

Signal detection analyses again indicated an absence of effects on either d' or $\log \beta$ for sex of participant, leading to the elimination of this design variable from the analyses to be reported. Mean values of d' and $\log \beta$ are presented in Figures 2 and 3. The results for d' indicated that participants were more sensitive to the fame variation for unfamiliar than for familiar names, $F(1, 37) = 14.78, p < .0005$, with no other effects significant. The results for $\log \beta$ showed, as before, that participants more readily assigned fame to familiar names compared with unfa-

miliar names, $F(1, 37) = 33.89, p < .0005$. Of greatest relevance to the assessment of implicit stereotyping, analysis of $\log \beta$ also showed, as before, that participants attributed fame more readily to male than to female names, $F(1, 37) = 5.93, p = .02$. The interaction effect on $\log \beta$ was nonsignificant, $F(1, 37) = 1.29$. However, the pattern of data replicated previous experiments, with the comparison of familiarized male and female names replicating the difference favoring male names, $t(37) = 2.62, p = .01$, and the unfamiliarized names showing no gender bias, $t(37) = .84, p = .41$.

Correlation Between Explicit Beliefs About Gender Equality and Implicit Fame Stereotyping

The eight items were scored so that higher scores indicated greater sexism or weaker beliefs in gender equality. Cronbach's coefficient alpha on standardized variables was .81. On the 5-point scale, the mean sexism score was 2.36 (minimum = 1.11, maximum = 3.78). The mean sexism score for male participants was 3.03 ($SD = 0.63$) and for female subjects was 2.33 ($SD = 0.56$), $t(36) = 3.55, p = .001$. In response to Item 9, which asked if participants endorsed the label *feminist*, 23% of the male participants did so (3 of 13), compared with 40% of the female participants (10 of 25). The presence of a gender difference on this most explicit measure of sexism and the absence of a gender difference in actual assignment of fame to male and female names shows the lack of a relationship between general explicit measures of gender egalitarianism and more specific implicit measures of gender stereotyping.

We computed a Pearson product-moment correlation to examine the relationship of the direct measure of sexism with each participant's d' and $\log \beta$ scores, separately for male and female names. There was no statistically significant relationship between the explicit measure of sexism and $\log \beta$ for either male names ($r = .13$) or female names ($r = .09$). In other words, this measure of explicit beliefs about gender equality did not predict variations in the criterion used to assign judgments of fame to male and female names. For d' , however, correlations with the explicit measure of sexism were unexpectedly and significantly negative (for male names $r = .34, p = .03$; for female names $r = .42, p = .02$). That is, the higher the sexism score, the lower the sensitivity to fame in general.

It was of greater a priori interest to examine the correlations of the explicit sexism score with (a) the male-name minus fe-

⁶ The items were: (1) When women marry, they should not change their names; (2) A woman should be the primary caregiver for her child; (3) Affirmative action programs for women are a necessity to alleviate the bias against women in the workplace; (4) Abortion is justified only in cases of rape, incest, or threat to the mother's life; (5) The typical American college education suffers from a glaring male bias; (6) Existing laws provide adequate protection for the rights of women; we do not need the Equal Rights Amendment; (7) Gender-specific language promotes sexism; such words are best eliminated from the language; and (8) Maintaining gender roles makes a relationship better. Two additional items were: (9) I am a feminist (Yes/No) and, if the answer to (9) was No, response to a follow-up item was requested: (10) I would not call myself a feminist, but I believe firmly in the equality of women and men (Yes/No).

male-name difference in sensitivity to fame (d') and (b) the female-name minus male-name difference in criterion for judging fame ($\log \beta$). If the implicit and explicit measures are related, the expected results are positive correlations—that is, the more sexist the participant, the greater the sensitivity to male than female fame and the higher the criterion for judging female fame relative to male fame. These correlations were computed using the averaged differences for the two pairs of d' and $\log \beta$ measures that were available for each participant—that is, averaging across familiar and unfamiliar name conditions. The correlation between the the explicit measure and the difference score on d' (the implicit measure) was .045, with 95% confidence interval $-.323 < r < .364$. Likewise, the correlation between the female–male difference score on the explicit measure and $\log \beta$ was $-.096$, with a 95% confidence interval of $-.234 < r < .401$.

Discussion

Like the two preceding experiments, Experiment 3 demonstrated implicit gender stereotyping in fame judgments. Fame was more readily conferred on male than female names, and this difference was obtained more strongly for familiarized names than for unfamiliarized names. This replication provides support again for the thesis that knowledge about the differential association of fame with the categories male and female results in the implicit use of a differential criterion in judging individual members of gender categories.

The unpredicted negative correlation between the experiment's explicit measure of sexism and d' scores indicated that participants with higher sexism scores were less sensitive to fame variation of names, that is, they were less knowledgeable about famous people in general. This finding of a relationship between general knowledge and implicit sexism deserves attention in future research. The finding of most relevance here, however, is the lack of relationship between expressed explicit beliefs about gender equality and the criterion for fame assignment. Participants who scored high on sexism did not systematically differ from those who scored low in their difference in criteria for judging male versus female fame. This finding is consistent with other findings of dissociation between implicit stereotypes and explicit attitudes toward race (e.g., Devine, 1989). However, it also is possible that the lack of a relationship between the explicit and implicit measures may be the result of an insensitive explicit measure of sexism. In particular, the explicit measure was constructed for the purpose of this experiment, and although it scored well on face validity and internal consistency, the lack of validation of these items cannot be dismissed. Thus, in Experiment 4 we included a more standard measure of explicit gender stereotypes to again perform a test of the relationship between implicit and explicit gender stereotyping.

Experiment 4: An Attempt to Eliminate the Implicit Stereotyping Effect

Experiment 4 had three aims. First, we introduced a manipulation to make the gender variable in the name stimuli more salient, although not explicit, to test the thesis that making the cue more salient might suffice to reduce implicit gender bias.

Second, we introduced a new direct measure of gender stereotypes that was different from the one used in Experiment 3, to further examine the relationship between explicit and implicit gender stereotyping. Experiment 3 may not have produced a relationship between the implicit and explicit measures because (a) the items were not derived from a validated scale measuring gender-related beliefs, and (b) the items may have measured gender beliefs that are not specifically related to gender beliefs that underlie fame judgments. Hence, we introduced a more traditional measure of explicit gender beliefs by asking participants to rate men and women (as social groups) on a variety of traits. Some of the selected traits reflected abilities that may be considered to be associated with fame (e.g., competent, ambitious, assertive). Third, Experiment 4 included an open-ended measure of participants' speculation about the purpose of the experiment and the strategies used in making fame judgments to provide a more thorough examination of participants' conscious appraisal of their judgment task than had been attempted in the preceding experiments. This allowed us to use a measure sometimes used in research on implicit memory when similar issues concerning the involvement of awareness arise (see Bowers & Schacter, 1990). Such measures allow elimination of the threat that conscious awareness of the source of influence may produce the effect and allow examination of differences between aware and unaware participants.

Method

Participants and Design

71 students enrolled in an introductory psychology course (38 women, 33 men) participated in the experiment, in two sessions 48 hr apart. For both sessions, participants were tested in their classroom. The design was a 2 (name gender: female or male) \times 2 (name exposure: familiar or unfamiliar) \times 2 (gender cue salience: high or low) \times 2 (participant sex: female or male) factorial, with gender cue salience and participant sex being between-subjects variables.

Materials

The lists of famous and nonfamous names were the same as those used in Experiments 2 and 3. We devised a new measure of explicit beliefs about gender for use in this experiment. Eighteen traits were selected from the Bem Sex Role Inventory (Bem, 1974) such that (a) 6 were considered to be stereotypically associated with men more than with women (assertive, competitive, self-sufficient, dominant, ambitious, individualistic), (b) another 6 were considered to be stereotypically associated with women more than with men (sympathetic, shy, warm, gullible, gentle, flatterable), and (c) 6 were epicene (truthful, conceited, reliable, solemn, adaptable, efficient). Instructions requested participants to rate both men and women on each trait, using a 7-point scale, with 1 = low, 4 = medium, and 7 = high association to the trait being rated. The instructions, which were designed to overcome college students' unwillingness to express stereotypes or even to participate in exercises that request them to do so, were as follows:

You will see a list of adjectives often used to describe people or groups of people. For instance, we know that men are (on average) taller than women and older people are (on average) more prone to certain ailments than younger people, etc. It is almost never the case that all members of a group share an attribute (for example, it is not the case that all men are taller than all women), but people

can easily make judgments about which group has more or less of some quality.

We are interested in your judgment of how typical some traits are of men and women. We are not interested in your judgment of yourself or your friends. *We would like you to base your judgment on your knowledge of the population of males and females.* Also, we are not interested in your judgment of what *ought* to be the case (i.e., whether one group ought to be more or less hardworking). Nor are we interested in your opinion about why these differences emerge. Instead, we are simply interested in your judgment about the strength of the relationship between these *personality attributes* and *gender*, as you believe them to exist currently.

Procedure

Approximately half of the male and female participants were assigned to the nonsalient-gender-cue condition ($n = 36$), and the other half were assigned to a salient-gender-cue condition ($n = 35$). The familiarization (prior exposure) and fame judgment procedures were identical to those of Experiments 2 and 3, with two exceptions. First, all participants made the pronunciation judgment in Session 1 on a 3-point scale rather than a 5-point scale, allowing the task to be completed more rapidly and without sacrificing the quality of name familiarization. Second, participants in the salient-cue condition were (a) given instructions that repeatedly referred to the stimuli as "male and female names" (rather than simply as "names") and (b) received stimulus forms in which the names were printed in gender-exclusive columns (half the participants in this condition received male names in the left column and female names in the right column, whereas the other half got the reverse order). Participants assigned to the nonsalient-cue-condition received the same instructions as in previous experiments.

In Session 2, after the fame judgment task, all participants received the 18-trait direct measure of gender stereotypes (described above). After completing all tasks, participants were asked to provide information about their awareness of purposes of the experiment. Participants were asked to indicate if they had any hypotheses about the purpose of the study, and if so, to briefly describe the hypotheses.

Results

As in previous experiments, we classified data from the fame judgment task into four within-subjects categories (familiar male, familiar female, unfamiliar male, unfamiliar female) and reduced them to hit and false alarm rates. As Figure 1 shows, hit rates were higher for familiar than unfamiliar names, $F(1, 70) = 30.59, p < .0005$; and for male names than female names, $F(1, 70) = 4.65, p = .034$; with a nonsignificant interaction effect, $F(1, 70) = 1.81$. False alarm rates were higher for familiar names than unfamiliar names, $F(1, 70) = 32.97, p < .0005$, and for male names than female names, $F(1, 70) = 5.82, p = .018$. As in Experiment 2, the interaction effect of name familiarity and gender took the form of especially high false alarm rates for familiar male names, $F(1, 70) = 6.82, p = .011$.

An analysis that examined the effect of the instruction to increase gender salience of names found no significant effect of instruction. The average of the d' scores across the four name conditions was 2.30 in the gender-salient instruction condition and 2.38 in the gender-nonsalient instruction condition, $F(1, 69) = .25, ns$. The average of the $\log \beta$ scores across the four name conditions was 1.88 in the salient-gender-instruction condition and 1.75 in the nonsalient condition, $F(1, 69) = .14, ns$. Consistent with the results of all previous experiments, signal

detection analyses again revealed an absence of effects associated with male versus female participants. The variables of subject sex and gender salience of instructions were, accordingly, dropped from subsequent analyses.

Mean values of d' and $\log \beta$ appear in Figures 2 and 3, respectively. For analyses of d' , there was a significant main effect of familiar versus unfamiliar names, $F(1, 70) = 7.24, p = .009$, with participants being more sensitive to unfamiliar than familiar names. There was no significant effect on d' of female versus male names, $F(1, 70) = 0.70, ns$, nor of the interaction of name familiarity and name gender, $F(1, 70) = 1.10, ns$. The results for $\log \beta$ showed for the fourth time that participants more readily assigned fame to familiar than unfamiliar names, $F(1, 70) = 27.99, p < .0005$, and that participants used a lower criterion when judging fame for male than female names, $F(1, 70) = 7.79, p = .007$. The interaction of name familiarity and name gender was also significant, $F(1, 70) = 7.99, p = .006$. Once again, this interaction effect took the form of an especially low criterion for the condition of familiar male names. A comparison of familiar male versus female names indicated, as in all previous experiments, a bias favoring male names, $t(71) = 3.97, p = .0002$, but no such difference for unfamiliar names, $t(71) = .40, p = .68$.

Postexperimental Questions

On the postexperimental questionnaire, we asked participants to generate hypotheses about the purpose of the study. The data from these postexperimental questionnaires were coded as follows by an assistant who was not familiar with the hypotheses of the experiment: A = no hypothesis provided, B = a hypothesis that previously exposed names would be treated differently from nonexposed ones, C = a hypothesis that male names would be treated differently from female names, D = hypotheses concerning effects of both previous exposure and name gender, and E = incorrect hypotheses. The data indicated that most participants were not aware of the intended purposes of the study. The data from 88% of the 71 subjects were categorized as A ($n = 33$) or E ($n = 24$), indicating lack of knowledge of hypotheses. The most commonly generated incorrect hypothesis guessed at some effect of Session 1's pronunciation task on Session 2's fame judgments. Of the remaining 14 participants, 12 suspected a relationship between the previous name exposure and fame judgment, although not all of these specified a direction of expected relationship between name familiarity and judgment.⁷ Of the 71 participants, only 2 mentioned that the gender of the names may have been of interest. Of these, 1 hypothesized that memory for male and female names may have been at issue, whereas the other was closer to penetrating the true purpose of the experiment: "I assume the pronuncia-

⁷ Participants guessing at a relation between the tasks is not surprising, because unlike other implicit stereotyping experiments (Banaji et al., 1993), no effort was made in these experiments to keep the two tasks unrelated in the participants' experience. In fact, the instructions for the fame task mentioned that some of the names had appeared on the previous list. It is therefore surprising that only 12 participants hypothesized a possible link between the exposure of names in Session 1 and the fame judgment task in Session 2.

tion thing was a red herring and this study is somehow gender based." No participants correctly produced hypotheses concerning both independent variables. We conducted an additional analysis, dropping the 14 participants who correctly mentioned one or both independent variables. None of the findings shifted from being significant to nonsignificant or vice versa, although magnitudes of test statistics tended to diminish with the smaller resulting sample size.

Correlations With Explicit Stereotype Measures

Traits were judged such that the higher the score on the 7-point scale, the greater the perceived strength between the trait and the social group "women" or "men." We derived difference scores from participants' ratings of the groups "men" and "women" by subtracting their ratings of men from ratings of women on each trait. Three final difference scores were produced, averaging across stereotypic male traits, stereotypic female traits, and neutral traits. The assumption here is that the greater the distance between the scores in the direction predicted by the stereotype (a positive difference for female-stereotypic traits and a negative difference for male-stereotypic traits), the greater the magnitude of explicit stereotyping.

First some descriptive statistics are provided. As expected, participants endorsed existing gender stereotypes of personality. Both male and female participants endorsed culturally prescribed explicit gender stereotypes. On the 7-point scale, mean ratings over all participants were as follows. For male-associated, female-associated, and neutral traits, respectively, the target group men was rated on average 5.38 ($SD = 0.74$), 3.67 ($SD = 0.70$), and 4.23 ($SD = 0.46$). For the target group women the corresponding averages were 4.06 ($SD = 0.71$), 5.21 ($SD = 0.70$), and 4.10 ($SD = 0.43$). The mean difference in the application of male traits to women minus men was 1.31 ($SD = 1.0$). For male participants this difference was 1.15 ($SD = 0.44$), and for female participants it was 1.36 ($SD = 1.14$), $t(69) = 1.28$, *ns*. Similarly, the overall mean difference in the application of female traits to women minus men was 1.54 ($SD = 1.16$). For male participants this difference was 1.36 ($SD = 1.14$), and for female participants it was 1.68 ($SD = 1.16$), $t(69) = 1.15$, *ns*. Thus, male and female participants showed equivalent evidence of explicit gender stereotypes.

As in Experiment 3, we computed product-moment correlations to examine the relationship between the female minus male difference scores on explicit stereotypes and the corresponding difference scores on the implicit measures of d' and $\log \beta$ (the latter scored as female minus male to keep expected direction of correlations positive). Also as in Experiment 3, the correlation was nonsignificant for $\log \beta$, $r = .10$, *ns*, but was significant for d' and was in the expected direction, $r = .31$, $p = .01$.

Discussion

Experiment 4 provided a fourth replication of the gender bias in fame judgments. Fame was more easily conferred on male names than on female names, and this effect was more apparent for familiarized names than for unfamiliarized names. In Experiment 4 we attempted to reduce the gender-related fame bias

by increasing the cue salience of name gender. However, the manipulation had no discernible effect. That is, alerting participants to the existence of male and female names in the stimulus set and organizing names into gender exclusive sections did not suffice to cause participants to be more egalitarian in their fame judgments. More specifically, it did not reduce the discrepancy in decision criteria (β) used to assign fame to male versus female names. A manipulation check to test whether the instruction produced the desired effect could not be included, because the gender-alerting instruction was itself supposed to implicitly prime the use of gender. In fact, the postexperimental question testing for awareness of gender indicated that participants in both instruction conditions were equally unaware that gender of the name could have influenced the judgment. Among the possible reasons for the ineffectiveness of the category-salience manipulation is the possibility that using a stricter criterion of judging female fame is not considered personally or socially reprehensible. However, such interpretations must await further research.

General Discussion

Summary of Findings

Figure 4 presents a comprehensive overview of the present findings, representing the three independent variables that were of theoretical interest—name gender, familiarization, and participant sex. Preparation of Figure 4 was preceded by conducting analyses of variance of the d' and $\log \beta$ measures from all four experiments, using both replication and participant sex as between-subjects variables. Figure 4's combination of results across experiments was justified by finding that replication was a rather unimportant source of variation. Of 16 statistical tests involving the replication variable for the d' and $\log \beta$ measures, none reached a Bonferroni-adjusted $\alpha = .05$ criterion of statistical significance. Only 1 of the 16 tests came close to that criterion. This was the interaction effect of name gender with replication; its near significance is associated with the discrepant finding of a name gender main effect on d' in Experiment 2, although the other three experiments yielded no such effect.

Effects of Participant Sex

Of special interest in the overall analysis were effects involving participant sex. Because of a priori interest in these effects, coupled with the substantial power of the experiments, it was judged appropriate to test them with unadjusted significance tests. Only one of eight effects—ignoring the nonsignificant effects involving replication—was statistically significant. This effect, which would have been significant also with a Bonferroni-adjusted criterion, was the interaction of participant sex and gender of names on d' , showing that men were more sensitive than women to actual fame of male names, whereas women were more sensitive than men to actual fame of female names, $F(1, 178) = 10.01$, $p = .002$. This outgroup homogeneity effect can be seen in the middle panel of Figure 4, where the heights of the male-name d' bars are greater for male than female participants, whereas the reverse is true for the female-name d' bars. However, although this interaction effect was significant,

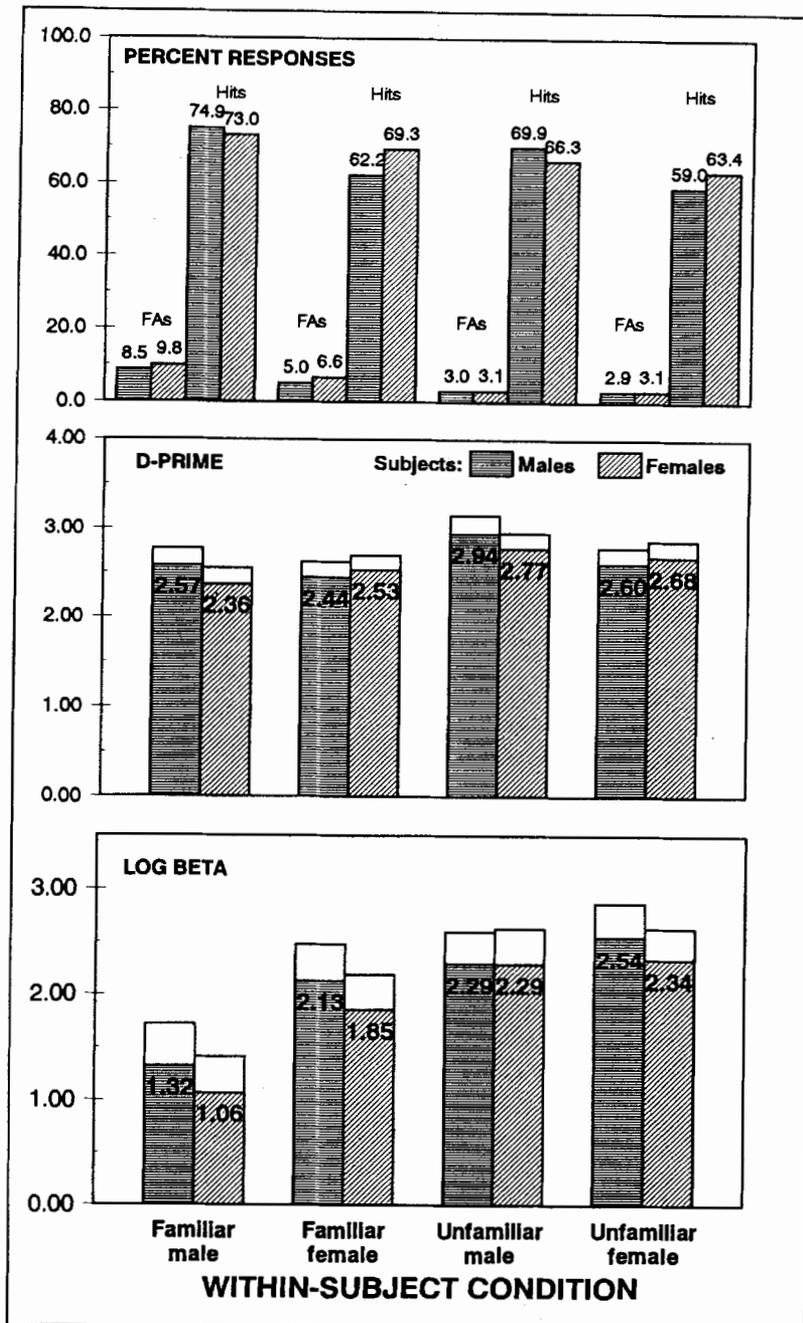


Figure 4. Mean hit and false alarm (FA) rates, d' , and $\log \beta$ for combined data of four experiments, showing upper halves of the 95% confidence interval for means on the latter two measures.

neither of its component contrasts was individually significant, for male names $t(184) = 1.53$, *ns*, for male versus female participants, and for female names $t(184) = 0.75$, *ns*. Other than this not especially surprising effect, there was no effect of participant sex on either d' or $\log \beta$ measure for which the F ratio statistic exceeded 1.0. Given the power of the combined data of the four experiments to detect small effects, the lack of participant-sex differences on $\log \beta$ (which was the major indicator of implicit

stereotyping in the present research) strongly indicates that implicit gender stereotypes are very similar for men and women, at least within the student populations sampled in the present research.

Effects on d'

In the highly powerful combined analyses, there were significant effects of name gender, prior familiarization, and their in-

teraction on d' , $F_s(1, 178) = 6.18, p = .014; 28.78, p < .0005$; and $8.34, p = .004$, respectively. These effects may be most easily understood as consequences of the strong effect of familiarization in increasing false alarm rates. This increase in false alarm rates had an overall effect of reducing sensitivity to the fame variation (producing the main effect of familiarization on d'), but did so especially for male names, for which the effect of familiarization on false alarms was especially great (producing the interaction effect on d'). The main effect of name gender on d' , superimposed on the interaction effect, is apparent with unfamiliarized names, $t(185) = 3.86, p < .0005$ (the difference for familiarized names was nonsignificant, $t[185] = 0.52$). This effect for unfamiliarized names indicates that, despite our effort to equate objective fame, the famous male names used in the research were better known to participants than were the famous female names. For this reason, although an interaction effect of name gender and familiarization was obtained on d' , an interpretation of the gender bias being located in sensitivity must await further evidence that can test this particular hypothesis more effectively.

Effects on Log β

Criterion for judgment of fame was powerfully influenced by the experimental manipulations, as can be seen in the bottom panel of Figure 4. There were significant effects of name gender, familiarization, and their interaction, $F_s(1, 178) = 43.81, 84.40, \text{ and } 13.96$, respectively, all $p < .0005$. These effects are readily apparent in Figure 4 as a lower value of $\log \beta$ in judgments of fame for male names relative to female names, for familiarized names relative to unfamiliarized names, and especially for familiarized male names (left pair of bars in bottom panel of Figure 4) relative to all other categories. The most compact description of the significant interaction effect is to note that the criterion difference in judging fame for male versus female names was significant for familiarized names, $t(185) = 7.16, p < .0005$, but not for unfamiliarized names, $t(185) = 1.44, ns$. This aspect of the interaction effect indicates the subtlety of the implicit stereotyping effect demonstrated in the present research. Even though there was a significant and strong main effect of male versus female names, that effect was carried almost entirely by familiarized rather than unfamiliarized names. Figure 4 also allows observation of the lack of difference between male and female participants in the implicit stereotyping effect. The criterion ($\log \beta$) difference on familiarized female names versus male names was $0.81 (= 2.13 - 1.32)$ for male participants, and a very similar $0.79 (= 1.85 - 1.06)$ for female participants. This nonsignificant difference between these differences, $t(184) = 0.11, p = .91$, reinforces the previous observation that implicit gender stereotype effects were as apparent for women as for men.

Theoretical Interpretation in Terms of Implicit Stereotyping

We offered a general definition of implicit stereotyping in the introduction as the introspectively unidentified traces of past experiences that mediate judgments of members of social categories (Greenwald & Banaji, 1995). Now, we clarify the condi-

tions under which the present false-fame procedure may be said to reflect implicit stereotyping. In these experiments, we capitalized on a widely held belief about the imbalanced association between fame and men, as compared with women. In this sense, the stereotyped association of gender and fame is not implicit. However, the use of such a belief in making fame judgments is assumed to be implicit. We report evidence supporting this assumption in Experiment 4, in which we found that participants did not report awareness of the influence of gender on fame judgment. Whether such a gender bias in false fame is obtained when participants are aware that the gender of the name may influence their judgment is an empirical question that we begin to explore in these experiments (Experiments 3 and 4), but this complex issue cannot be fully addressed in the present article, which reports the first experiments documenting this bias.

These experiments contribute to theory in more than one way. First, they challenge the generalizability of existing theories of implicit memory. Implicit memory effects that involve social stimuli (as do the fame experiments in the form of male and female names) may not be obtained (or may be only weakly obtained) if the social properties of the target (gender) and the dimension of judgment (fame) are not strongly associated. The main finding of these experiments is that the social category of the target moderates the influence of repetition priming. The importance of social category in priming effects is seen in these experiments in the significantly greater repetition priming effect for targets that are equated in exposure but vary in social category. Existing theories of priming must be modified to account for these data by introducing the potentially influential dimension of social beliefs.

These experiments add to a growing body of research on the implicit or unconscious ways in which stereotyping effects may be produced. In particular, these findings point out the potential to obtain relatively robust effects of stereotyping when participants (irrespective of their own social category and explicit beliefs) are called on to make judgments that do not explicitly draw attention to their own conscious beliefs about gender egalitarianism and the purpose of the social variation of the names. This finding challenges theories of stereotyping that assume that differences in explicit awareness of stereotypes can predict implicit stereotyping.

These experiments also support the proposition that experiences that might influence a judgment actually do so only when the judgment is inferentially appropriate in the situation. One basis for a judgment being inferentially appropriate is that it agrees with the stereotype of a social category that is salient in the situation. For example, Banaji et al. (1993) found that primes designed to activate judgments of aggressiveness did so when a male stimulus person was being judged but not when a female was judged. Likewise, primes designed to activate judgments of dependence did so when a female stimulus person was judged but not when a male was judged. These effects presumably depended on the stereotypical association of aggressiveness with males and of dependence with females. In the present experiments, the judgment of fame was inferentially more appropriate for a familiar-sounding name that was male than for one that was female.

Experiments 3 and 4 showed that individual differences in explicit beliefs about gender equality and gender stereotypes were unrelated to individual differences in stereotypes that were expressed in fame judgments. This lack of a relationship is consistent with other findings in the stereotyping literature and with the theoretical position that implicit and explicit stereotypes may operate independently of each other. However, these results should be viewed as tentative evidence for the dissociation between explicit and implicit stereotypes. Evidence of the lack of a relationship between explicit and implicit stereotypes would gain in predictive power if research on implicit stereotyping simultaneously obtained predicted correlations with other implicit measures. Identifying the predictive validity of implicit stereotyping measures will be among the most important issues facing the next wave of research on implicit stereotyping and prejudice.

Use of Signal Detection Analysis

The expression of stereotypes in social judgments has been established empirically in many contexts. As noted in the introduction, these effects might consist of some mixture of (a) diminished sensitivity to differences in possession of a stereotyped trait among members of a group with whom the trait is stereotypically associated, or (b) bias in the form of overattribution of the stereotyped trait to members of the group. Signal detection methods are potentially valuable in the study of stereotypes because they can distinguish sensitivity from bias (criterion) as stereotype components and also permit finding evidence for both types of effects in the same situation.

The present research found evidence for both sensitivity and bias effects associated with gender stereotypes that were expressed in fame judgments. The effects of gender on sensitivity to fame were complicated by the difficulty of selecting male and female names that were objectively equal in actual fame. Nevertheless, the familiarization procedure reduced sensitivity to fame and did so more for male names than for female names. Stronger and more readily interpretable effects occurred in the form of bias to associate fame with men more than women. The bias was especially apparent when participants were judging names that had been given a boost in familiarity by participants' brief encounter with them in Session 1 of the experiment.

Using Judgments of Names to Investigate Implicit Stereotyping

The differential attribution of fame in these experiments is assumed to have occurred because of the co-occurrence of two conditions. First, the sought judgment of fame was systematically associated with one social group more than another (men more than women), and second, participants were placed in a specific condition of uncertainty created by an inability to accurately identify the source of name familiarity. Identifying these properties suggests that such implicit stereotyping effects might be apparent with attributes other than fame and with social categories other than male and female. For example, the attribute "criminal" or "athlete" may be more easily assigned to Tyrone Washington than to Edwin Washington, and the professional role of "opera singer" more easily attributed to Pauline

Manchetti than Pauline Jones. The procedure of obtaining judgments of names and its signal detection analysis are potentially usable for diagnosing implicit stereotypes associated with a wide variety of social categories that are encoded in names.

The use of names as stimuli in social psychological research has substantial precedents. Allport (1954) reported the use of names in an experiment by Wax (1948) in which letters were written to resorts and hotels requesting room reservations for the same dates under two names, Mr. Greenberg and Mr. Lockwood. The results were striking: "Mr. Lockwood" received responses from 95% of the resorts, and 93% offered accommodation. "Mr. Greenberg" received responses from 52% of the resorts, and 36% offered accommodation. Since then, scores of experiments have used names to indirectly invoke social category (see Goldberg, 1968, for a name-based procedure that has since been widely used; see Swim, Borgida, Maruyama, & Myers, 1989, for a meta-analysis of such studies; see Kasof, 1993, for a discussion of interpretational confounds of name-based experiments). In the present experiments, as in many daily interactions, perceivers may have had little objective knowledge about a person they were called on to judge. In this knowledge vacuum, the social category information conveyed by a name may assume great significance. In addition to signaling gender very readily, names often also indicate race or ethnicity, nationality, religion, class or caste, and age. Names may be especially useful as research stimuli because they effectively communicate information about social categories without calling explicit attention to that fact.

Although the hypothesis that suggests the involvement of implicit memory for names in producing implicit stereotyping is unique to these experiments, the underlying ideas are common to previous theories of the role of cognition in stereotyping. For example, the assumptions underlying these experiments are in agreement with Rothbart (1981), who said: "I would argue that the need for action demands that we choose one response alternative over another, and that demand often requires the use of data that may be quantitatively and qualitatively deficient. No rational person would choose a spouse solely on the basis of his or her first name ([Oscar] Wilde's play notwithstanding), but if we had no other information about a potential spouse except a first name, wouldn't there be agreement that Eric and Samantha are more exciting than Bill and Jane?" (p. 178).

It is not expected that the observed gender stereotyping effect is a function of the selection of particular first names to represent gender. In Experiment 1, we used 288 nonfamous names (144 male, 144 female) and 144 famous names (72 male, 72 female), ruling out the potential confound that the gender stereotyping effect may have been produced by a small set of unmatched male and female names. Although the argument offered by Kasof (1993) that the effects of some gender stereotyping studies may have been produced by the unwelcome covariation of male and female names with other attributes (such as competence or attractiveness) has merit in the context of the specific studies discussed in that article, such an argument cannot easily hold for studies in which large samples of male and female names are used.

Everyday social interactions provide countless conditions of uncertainty (and, as in this case, uncertainty produced by fa-

miliarity) about the cause of familiarity under which judgments about people are required and provided. These results suggest that under such conditions, perceivers use social category information (elicited by even seemingly superficial markers such as a name) by using criteria of differing strictness in judging individual members. The contribution of the present findings is in locating such a difference in the criterion used in judgment as the basis of stereotyping.

Conclusions and Implications

The main empirical discovery of the present experiments is that a lower criterion is used in fame judgments of familiar men than of women. This benefit to men is assumed to be unconscious and produced by an interaction of two factors: (a) an inability to distinguish two sources of familiarity (actual fame vs. fluency from previous exposure) and (b) the use of social category knowledge in individual judgment. The main methodological contributions of this article are (a) the introduction of signal detection analyses to isolate the criterion component of implicit stereotyping and (b) introduction of the name-judgment procedure as one with potential broad applicability for examining implicit cognition in stereotyping. The discovery of such implicit stereotyping effects bolsters the recommendation for a departure in thinking about individual responsibility in the everyday perpetration of stereotyping. In particular, such discoveries seriously question the wisdom of assuming individual responsibility and extracting individual punishment for unconscious acts that may occur outside the social perceiver's awareness. Such findings render moot the possibility of spontaneous, individually devised compensatory strategies. Rather, efforts must focus on alternative strategies to (a) generate procedures by which individuals can routinely recognize the effects of implicit stereotyping and (b) modify environments such that the operation of implicit stereotypes is inhibited.

The pernicious consequences of persistent implicit stereotyping for the social target, however, raise questions about the availability of adequate measures to safeguard against the occurrence of such actions. In particular, they raise the issue of target protection against judgments of which the target is unaware and which therefore render moot standard legal procedures such as due process. Such thinking has led us to discuss in another article the relative merits of blinding, consciousness raising, and affirmative action procedures (Greenwald & Banaji, 1995). Although adequate solutions to such issues lie in the future, discovering the cognitive conditions that produce implicit stereotyping can inform about the operations of social knowledge and contribute to proposals for social engineering.

References

- Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.
- Aronson, E. (1988). *The social animal*. New York: Freeman.
- Banaji, M. R., & Greenwald, A. G. (1991, June). *Measuring unconscious attitudes*. Paper presented in a symposium on the psychological unconscious at the third annual meetings of the American Psychological Society, Washington, DC.
- Banaji, M. R., & Greenwald, A. G. (1994). Implicit stereotyping and unconscious prejudice. In M. P. Zanna & J. M. Olson (Eds.), *The psychology of prejudice: The Ontario Symposium* (Vol. 7, pp. 55-76). Hillsdale, NJ: Erlbaum.
- Banaji, M. R., Hardin, C., & Rothman, A. J. (1993). Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology*, 65, 272-281.
- Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1-40). Hillsdale, NJ: Erlbaum.
- Basow, S. (1986). *Gender stereotypes: Traditions and alternatives*. Monterey, CA: Brooks/Cole.
- Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, 42, 155-162.
- Bowers, J. S., & Schacter, D. L. (1990). Implicit memory and test awareness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 404-416.
- Broverman, I., Vogel, S. R., Broverman, D. M., Clarkson, F., & Rosenkrantz, P. S. (1972). Sex role stereotypes: A current appraisal. *Journal of Sex Roles*, 28, 59-78.
- Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of black and white discrimination and prejudice: A literature review. *Psychological Bulletin*, 87, 546-563.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5-18.
- Dovidio, J. F., Evans, N. E., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology*, 22, 22-37.
- Eagly, A. H., & Mladinic, A. (1989). Gender stereotypes and attitudes toward women and men. *Personality and Social Psychology Bulletin*, 15, 543-558.
- Fiske, S. T. (1989). Examining the role of intent: Toward understanding its role in stereotyping and prejudice. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 253-283). New York: Guilford Press.
- Gaertner, S. L. (1973). Helping behavior and discrimination among liberals and conservatives. *Journal of Personality and Social Psychology*, 25, 335-341.
- Geis, F. (1993). Self-fulfilling prophecies: A social psychological view of gender. In A. E. Beall & R. J. Sternberg (Eds.), *Perspectives on the psychology of gender* (pp. 9-54). New York: Guilford Press.
- Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, 60, 509-517.
- Goldberg, P. (1968). Are women prejudiced against women? *Transaction*, 5, 28-30.
- Graf, P., & Schacter, D. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 501-518.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Greenwald, A. G. (1990). What cognitive representations underlie social attitudes. *Bulletin of the Psychonomic Society*, 28, 254-260.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4-27.
- Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotype judgments. *Journal of Experimental Social Psychology*, 12, 392-407.
- Hewstone, M., Islam, M. R., & Judd, C. M. (1993). Models of crossed categorization and intergroup relations. *Journal of Personality and Social Psychology*, 64, 779-793.
- Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, 110, 306-340.

- Jacoby, L. L., Kelley, C. M., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology*, *56*, 326-338.
- Jacoby, L. L., Lindsay, D. S., & Toth, J. P. (1992). Unconscious influences revealed: Attention, awareness, and control. *American Psychologist*, *47*, 802-809.
- Jacoby, L. L., & Witherspoon, D. (1982). Remembering without awareness. *Canadian Journal of Psychology*, *36*, 300-324.
- Kasof, J. (1993). Sex bias in the naming of stimulus persons. *Psychological Bulletin*, *113*, 140-163.
- Katz, D., & Braly, K. (1935). Racial prejudice and racial stereotypes. *Journal of Abnormal and Social Psychology*, *30*, 175-193.
- Kolers, P. (1976). Reading a year later. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 554-565.
- Linville, P. W., Salovey, P., & Fischer, G. (1986). Stereotyping and perceived distributions of social characteristics: An application to in-group-outgroup perception. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 165-208). New York: Academic Press.
- McHenry, R. (1980). *Famous American women: A biographical dictionary from colonial times to the present*. New York: Dover.
- Ostrom, T. M., & Sedikides, C. (1992). Out-group homogeneity effects in natural and minimal groups. *Psychological Bulletin*, *112*, 536-552.
- Park, B., & Rothbart, M. (1982). Perception of out-group homogeneity and levels of social categorization: Memory for the subordinate attributes of in-group and out-group members. *Journal of Personality and Social Psychology*, *42*, 1051-1068.
- Perdue, C. W., & Gurtman, M. B. (1990). Evidence for the automaticity of ageism. *Journal of Experimental Social Psychology*, *26*, 199-216.
- Pratto, F., & Bargh, J. A. (1991). Stereotyping based on apparently individuating information: Trait and global components of sex stereotypes under attention overload. *Journal of Experimental Social Psychology*, *27*, 26-47.
- Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. *Annual Review of Psychology*, *39*, 474-543.
- Roediger, H. L. (1990). Implicit memory: Retention without awareness. *American Psychologist*, *45*, 1043-1056.
- Roediger, H. L., & Blaxton, T. A. (1987). Effects of varying modality, surface features, and retention interval on priming in word fragment completion. *Memory & Cognition*, *15*, 379-388.
- Roediger, H. L., Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between measures of retention: A processing account. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness* (pp. 3-42). Hillsdale, NJ: Erlbaum.
- 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., Evans, M., & Fulero, S. (1979). Recall of confirming events: Memory processes and the maintenance of social stereotypes. *Journal of Experimental Social Psychology*, *15*, 343-355.
- Ruble, D. N., & Ruble, T. L. (1982). Sex stereotypes. In A. G. Miller (Ed.), *In the eye of the beholder: Contemporary issues in stereotyping* (pp. 188-252). New York: Praeger.
- Schacter, D. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 432-444.
- Secord, P. (1959). Stereotyping and favorableness in the perception of Negro faces. *Journal of Abnormal and Social Psychology*, *59*, 309-315.
- Smith, E. R., Ferree, M. M., & Miller, F. D. (1975). A short scale of attitudes toward feminism. *Representative Research in Social Psychology*, *6*, 51-56.
- Spence, J. T., & Helmreich, R. (1972). The Attitudes Toward Women scale: An objective instrument to measure attitudes toward the rights and roles of women in contemporary society. *JSAS Catalog of Selected Documents in Psychology*, *2*, 66.
- Swim, J., Borgida, E., Maruyama, G., & Myers, D. G. (1989). Joan McKay versus John McKay: Do gender stereotypes bias evaluation? *Psychological Bulletin*, *105*, 409-429.
- Taylor, S. E. (1981). A categorization approach to stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 83-114). Hillsdale, NJ: Erlbaum.
- Tulving, E., & Schacter, D. (1990). Priming and human memory systems. *Science*, *267*, 301-306.
- Wax, S. L. (1948). A survey of restrictive advertising and discrimination by summer resorts in the Province of Ontario. *Canadian Jewish Congress: Information and Comment*, *7*, 10-13.
- Who's Who in America*. (1988). Wilmette, IL: R. R. Bowker.
- Williams, J. E., & Best, D. L. (1990). *Measuring sex stereotypes: A multinational study*. Newbury Park, CA: Sage.

Received September 23, 1993

Revision received April 28, 1994

Accepted April 29, 1994 ■