European Journal of Social Psychology, Vol. 24, 623–639 (1994)

An ethnic bias in scientific citations

ANTHONY G. GREENWALD and
ERIC S. SCHUH
University of Washington

Abstract

Recent experimental findings of subtle forms of prejudice prompted this search for a similar phenomenon outside the laboratory. In Study 1, with a sample of more than 12 000 citations by North American social scientists, names of both citing and cited authors were classified as Jewish, nonJewish, or other Author's name category was associated with 41 per cent greater odds of citing an author from the same name category Study 2 included over 17 000 citations from a much narrower research domain (prejudice research), and found a similar (40 per cent) surplus in odds of citing an author of the author's own ethnic name category. Further analyses failed to support two hypotheses — differential assortment of researchers by ethnicity to research topics, and selective citation of acquaintances' works — that were plausible alternatives to the hypothesis that the observed citation discrimination revealed implicit (unconsciously operating) prejudicial attitudes. Given the sociopolitically liberal reputation of social scientists (and of prejudice researchers especially), it seems unlikely that the observed bias in citations reflected conscious prejudicial attitudes.

INTRODUCTION

In North America, discrimination based on religious ethnic identity is thought to occur only in marginal, politically extreme groups. Perhaps surprisingly, the present research discovered such discrimination in a group of scientists who are commonly regarded as lacking in racial or ethnic bias[1] In particular, evidence for discrimination, based on Jewish versus nonJewish ethnicity of authors' names, was found in the scientific citations that are documented in the *Social Sciences Citation Index*.

Addressee for correspondence: A. G. Greenwald, Department of Psychology (NI-25), University of Washington, Seattle, WA 98195, U.S.A.

[1] The terms *discrimination* and *bias* are sometimes used with implication of prejudicial attitude, as in 'gender discrimination' and 'racial bias' However, these same two terms are often used descriptively, with no implication of prejudice, as in 'sensory discrimination' and 'response bias' In order to distinguish these two usages, this manuscript uses 'ethnic discrimination' only in the descriptive sense — that is, to designate differential behaviour toward ethnic groups, with no implied interpretation in terms of mediation by prejudicial attitude. By contrast, 'ethnic bias' is used to indicate conscious or unconscious attitudinal dispositions that may be responsible for observed discrimination.

CCC 0046–2772/94/060623–17

Received 28 January 1993
Accepted 15 August 1993

The authors' interest in possible ethnic discrimination in scientific citations was provoked by past findings of racial discrimination by people who are expected not to be prejudiced (e.g. Baxter, 1973; Clark and Clark, 1947; DeFleur and Westie, 1958; Devine, 1989; Gaertner, 1976; Linn, 1965). Such findings suggest that race prejudice is, in effect, at large in society and, at least in its subtler forms, may be apparent equally among those who avow it and those who disavow it.

The present research was also prompted by a theoretical analysis of *implicit attitudes*. In this analysis (Greenwald, 1990; Greenwald and Banaji, in press), an attitude toward one attribute of an object is said to operate implicitly when it influences an evaluative response to another attribute *and* the actor is unaware of this influence. The familiar halo effect (see Cooper, 1981; Eagly, Ashmore, Makhijani and Longo, 1991 for reviews) provides an example. In the case of the halo effect, the attitude toward one aspect of a target person (usually, the target's physical attractiveness) influences the evaluation of another aspect (such as an attractiveness-unrelated trait, or an essay written by the target person — e.g. Landy and Sigall, 1974), despite the evaluator's being unaware of this influence.

The present studies investigated the behaviour of scientists citing other scientists' publications. Most often, citing a published work conveys a positive evaluation of that work. Could this evaluation be influenced by the citing author's attitude toward some objectively irrelevant attribute of the work? The conception of implicit attitudes suggests that, indeed, citations might implicitly express attitudes toward other attributes of the authors of potentially citable works. The examination of citations is especially interesting because the actors involved — that is, scientist-authors — may lack introspective awareness of the attitudes that can be expressed implicitly by their citations. This situation contrasts with that of the typical halo effect study, in which the evaluator may be well aware of perceiving the target person as physically attractive or unattractive, even while not recognizing that this attitude influences other judgments.

In examining citations, the present research investigated possible discrimination based on Jewish versus nonJewish ethnicity of citing and cited authors. In principle, this research could have used any contrast of ethnic identities that is identifiable on the basis of surnames. The Jewish–nonJewish contrast was chosen because it was one of very few such contrasts for which the smaller category (Jewish surnames) occurred frequently enough in the available archives (citation indexes) to be useful for testing hypotheses.

Pilot study

A preliminary small study sought to determine the extent of name-Jewishness-based discrimination (if any) in scientific citations. A computer program generated random-number pairs that were used to sample 640 citations, equally divided among recent volumes of the *Science Citation Index*, the *Humanities Citation Index*, and the *Social Sciences Citation Index*. The first number in each pair was interpreted as a column number (each of several columns of citations per page has a unique ordinal number in each volume); the second number was interpreted as a distance in mm from the top of the chosen column. For each citation thus found using the 640 number pairs as coordinates, the *citing* author's name was the surname of the first author of the publication containing the citation, and the *cited* author's name was the surname

of the first author of the cited publication. For 1145 different surnames thus obtained (135 were duplicates), three judges independently rated, on 5-point scales and in ignorance of the sampling origin of the name, both the name's Jewishness and its likelihood of being North American (i.e., that the person so named was from the U.S. or Canada). The intercorrelations (Pearson r) of the three judges' ratings on the 5-point Jewishness scale ranged from 0.50 to 0.60, based on 525 names that were above the median (of the average of the three judges) on likelihood of being North American. Those below the median were omitted because the judges were less confident in judging them, and assigned mid-scale ratings to many of them.

The pilot data were analysed by computing the regression of rated Jewishness of cited author's name on rated Jewishness of citing author's name, excluding 30 self-citations that would have inflated the regression coefficient. A positive value for the regression coefficient indicates possible discrimination — that is, a positive relation between Jewishness of citing and cited authors' names. The obtained coefficient *was* positive, but was also weak and nonsignificant ($b = +0.044$, $F(1,608) = 1.16$, n.s.).

A 95 per cent confidence interval on the pilot study's observed regression coefficient ($-0.037 < b < +0.126$) provided an initial estimate of Jewish ethnicity-based discrimination in citations. These confidence-interval bounds suggested the existence of, at most, a weak relationship. However, a weak effect is potentially important in a population of millions of citations annually. (The potential importance of such small effects is considered further in the Discussion section.) If an effect of the size indicated by the observed regression coefficient ($b = +0.044$) is valid, even a sample of 10 000 randomly selected citations would provide an underpowered hypothesis test. Therefore, the main study not only increased the sample size greatly (by a factor of about 20), but also — in order to increase sensitivity to small effects — changed the pilot study's random sampling design to one using *selective sampling* of citing author names that were judged unambiguously Jewish or nonJewish[2] (The Discussion section, below, considers the statistical power advantage of this selective sampling strategy)

STUDY 1

Method

From the list of surnames gathered in the pilot study, the present authors identified ones that had been rated, in the pilot study, as both (a) highly likely to be North American (average rating greater than 4.5 on the 5-point scale) and (b) either very low (average lower than 1.5) or very high (average greater than 4.5) in likelihood of being Jewish. Twelve likely Jewish surnames (J) and 24 likely nonJewish surnames (NJ) were selected in this fashion. Some others that met the same criteria, but were not used in Study 1, were: NJ — Ballard, Chambers, Davenport, Erickson, Hawkins, McBride, Paine, Watkins, and Webster; J — Baum, Buchman, Cohn, Edelman, Fried, Goldstein, Seligman, Siegel, Silberstein.

[2] Name Jewishness/nonJewishness is, of course, only an uncertain indicator of actual Jewishness/nonJewishness. The present authors were aware of some citing authors whose names were misclassified by the procedures used in this series of studies. Because these errors could only reduce evidence of bias, they were allowed to stand.

From the *1987 Annual Source Index* volumes of the *Social Sciences Citation Index (SSCI)*, all articles published by all authors with the 36 selected surnames were examined. For each article published in the one-year period covered by the volume, the *SSCI Source Index* provides the name and institutional affiliation of the first author, the names of coauthors, and a complete list of articles that were cited in each article's reference list. Articles were retained for the study if the author's institutional affiliation was in the U.S. or Canada[3] The final sample included 3658 citations in articles by 134 J authors and 8479 citations in articles by 343 NJ authors. (Of these 627 were self-citations, 252 J and 375 NJ, all of which were excluded form major analyses.)

Photocopies were made of the *SSCI Source Index* listing of the complete set of citations from each selected article. Each of these lists was then prepared for rating by (a) obscuring the author's name both at the head of the listing and on any lines giving self-citations, (b) taping randomly ordered sets of clipped listings onto sheet, and (c) making enlarged photocopies of these sheets for use by judges. For a randomly selected half of the sheets, categorizations were made independently by two judges (these judgments constituted subsets 1A and 1B of data set 1). One of the judges then provided the same judgments for the remaining half of the sample (subset 1C). The judges categorized each cited name (omitting anonymous and corporate citations, as well as the obscured self-citations) in a two-stage process. First, names were categorized as North American or other; and second, names judged as North American were further classified as (a) J, (b) NJ, or (c) uncertain, using each judge's subjective criteria. Non-North American names were excluded because of the judges' lack of confidence in judging the J/NJ distinction for them (see footnote 3). Separately from these judgments of cited authors' name ethnicity, one judge determined the disciplinary identification of each author, using information provided by both the author's departmental affiliation and the title of the journal(s) in which the author had published. On the basis of these ratings, each author was assigned to one of 15 disciplines.

Measures and analyses

The criterion measure used in all data analyses was the ratio of NJ to J cited authors. It was computed simply as the total number of NJ citations for a given author, divided by that author's total number of J citations. The criterion measure was used in a ratio format because (unlike a subtractive difference measure) the ratio is logically independent of the [highly variable] number of citations in individual cases. Solutions to some technical problems that were associated with use of this *citation odds* index are given in the Appendix. The Appendix also describes two additional analysis strategies that were used with each data subset in order to assure that the main reported analyses did not produce idiosyncratic results. A logarithm transform of the citation odds index was used for all statistical tests, in order to correct for the expected (and observed) heavy skewing of untransformed ratio indexes.

The citation odds index was computed for each author. Because the index was based on a variable number of citations, ranging between 1 and 200 citations per

[3] Restriction to North American-resident authors minimized inclusion of European names, which were difficult for North American judges to code. Restriction to the *SSCI* served a similar purpose, because both the *Humanities Citation Index* and *Science Citation Index* contained substantially higher proportions of authors from outside of North America than did the *SSCI*.

author, it was decided to use a weighted analysis in which authors having larger numbers of citations were given greater weight. Accordingly, each case (i.e. author) was assigned a weight computed approximately (see Appendix) as the square root of the number of citations on which its index was based. In order not to allow the few very largest cases to dominate the analysis, no case was allowed to have a weight larger than that associated with 50 citations. This weighting scheme was, in some respects, intermediate between unweighted and standard weighting strategies (see Appendix), and will be referred to as an *attenuated weight* analysis. The interjudge reliability of the attenuated weight (log) citation odds index was 0.65, computed using a regression procedure that is described in the Appendix.

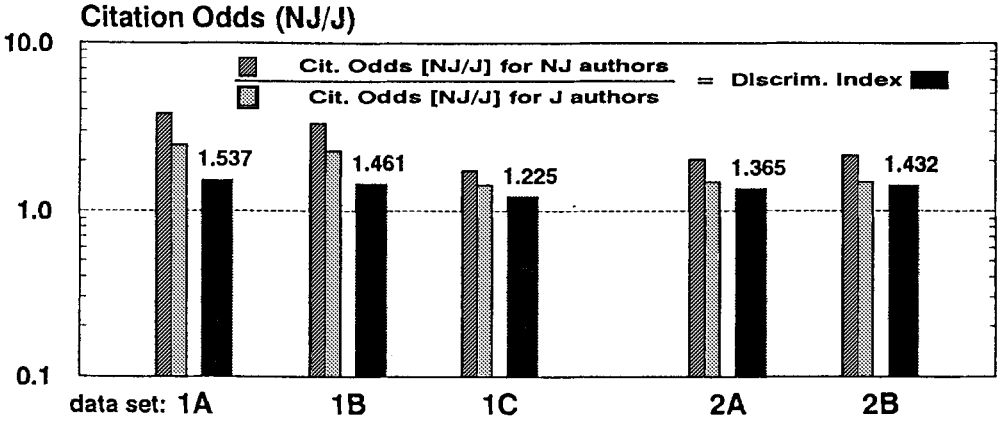

Figure 1. Citation discrimination. The first two bars in each group give the ratio of citations of works by nonJewish-named (NJ) authors to those by Jewish-named (J) authors, separately for NJ and J citing authors. Values of the discrimination index (solid bars) significantly exceeded 1.0 (= no difference in citation patterns for NJ and J authors) for all five data subsets in the two studies. All plotted values are based on geometric means of untransformed ratios — that is, the antilog of the mean of the log transforms — plotted on a logarithmic ordinate

Results

Figure 1 presents mean values for the (attenuated weight) citation odds measure, as a function of citing author's name ethnicity Values greater than 1.0 indicate more citation of NJ than J authors. (The index's values are *expected* to exceed 1.0 because NJ authors were in the majority) Name-ethnicity-based discrimination is indicated by higher values of the odds measure for NJ than for J citing authors. Such discrimination was apparent in each of the three subsets of Study 1, and is measured by Figure 1's *discrimination index*, which divides the mean odds ratio for NJ authors by that for J authors[4] Values of the discrimination index greater than 1.0 indicate discrimination in the direction of greater citation of own ethnicity. In all three analyses for Study 1, the index value was significantly greater than 1.0

[4] Note that the discrimination index is not a value that is available for individual authors, but is a single number for an entire data subset, based on the comparison of one summary number (the mean citation odds index for NJ authors) with another (the same mean index for J authors).

— that is, the citation odds ratio was significantly greater for NJ than for J citing authors. For subsets 1A, 1B, and 1C, respectively: $F(1,228) = 16.14$, $p < 0.0001$, $F(1,223) = 11.27$, $p = 0.0009$; and $F(1,243) = 4.88$, all $p = 0.028$, two-tailed. For the average across the three analyses of Study 1, the citation odds measure was 40.8 per cent greater for NJ than for J citing authors. (The method of computing these significance tests is described in the Appendix.)

Discussion

Although Study 1 demonstrated citation discrimination, its results tell nothing about who is discriminating. It is possible that only J authors are over-citing their J colleagues, or only NJ authors are over-citing their NJ colleagues, or that discrimination of a type resembling ingroup bias is occurring approximately equally in both groups. Additionally, the results tell nothing about the cause of the observed discrimination. The discrimination may be rooted in an ethnic prejudice. However, it may also be an artifact of processes that have nothing to do with ethnic bias or prejudice.

Because this study used archival data, there was no possibility of directly observing a mediating process. A conclusion that some form of ethnic bias is operating depends on considering and ruling out plausible artifactual alternative interpretations. One possible source of artifact is ethnic variation in disciplinary preferences. Assume that NJ researchers are attracted more to field 1 and J researchers more to field 2. As a consequence, researchers in field 1 should be both (a) more likely to have NJ names and (b) more likely to cite NJ authors than do researchers in field 2. Without any bias, this *differential assortment* could produce a pattern of citation discrimination resembling that shown in Figure 1. Of course, differential assortment could itself be a consequence of ethnic prejudice or bias. However, it could also be a benign consequence of cultural differences that are unrelated to prejudice.

In order to assess the possibility of explanation in terms of differential assortment, the 15 disciplines identified in Study 1 were grouped into four clusters such that the ratio of NJ cited authors to J cited authors was homogeneous within clusters. Significance tests for the effect of citing author's name ethnicity were then performed in a regression analysis in which disciplinary cluster (three dummy variables) was entered hierarchically prior to author's name ethnicity. Assessment of the differential assortment artifact was made in terms of the reduction in size of the regression coefficient for the author's ethnicity predictor, compared to the original analysis, which did not include disciplinary cluster as a predictor. Although the author-ethnicity effect was reduced from being significant (at $\alpha = 0.05$, two-tailed) to nonsignificant in subset 1C, still the reduction in regression coefficient for author's ethnicity averaged only 13.5 per cent across the three subsets, providing little evidence for a differential assortment artifact.

STUDY 2

Although Study 1 produced little or no evidence for a differential assortment artifact, its possibilities for detecting that artifact were quite limited. Study 2 sought to correct this limitation of Study 1, principally by using a design that reduced the plausibility of a differential assortment artifact. In order to do this, Study 2 used a topic boundary

that was narrow enough to make it unlikely that differential assortment could occur within that boundary Further, the specific topic chosen was one for which conscious prejudice by authors should be minimized: the sample was limited to prejudice researchers. (It is assumed that researchers who study prejudice are, for the most part, persons who have an explicit goal of attempting to reduce prejudice.) Lastly, Study 2 included data on geographic region and institutional location of authors, in addition to disciplinary affiliation, in order to allow even narrower identification of subgroups of researchers.

Method

A set of articles reporting research primarily on prejudice was selected by, first, identifying 14 'target' scientists whose works were prominently cited in publications about prejudice between 1981 and 1985. These 14 were chosen because they were authors of major single works on prejudice published prior to 1960, of theoretical analyses of attitude-behaviour relations that are frequently cited in the prejudice literature, or of specific influential recent (post-1969) publications, or because they had made sustained career contributions to research on prejudice.

The 1981–1985 five-year summary of the *SSCI Citation Index* listed 1712 articles that cited the prejudice-related works of one or more of the 14 target scientists. Surnames of the first authors of these 1712 articles were then categorized as J, NJ, or uncertain by two judges, as in Study 1. Of the 403 articles for which author names were classified identically as J or NJ by both judges, the 356 for which the author was located in the U.S. (82 per cent), Canada (5 per cent), or in an English-speaking institution in Europe or Australia (13 per cent) were retained for analysis. Excluding author self-citations (as in Study 1), all works cited in these 356 articles as listed in the 1981–1985 *SSCI Source Index*, were included in Study 2. The 82 retained J-authored publications (by 72 different authors) cited a total 3648 works, and the 274 NJ-authored articles (by 234 different authors) cited a total of 13 924 works. As in Study 1, after photocopying the *Source Index* listings for the 356 articles and obscuring author-identifying information, the 17 572 names of cited authors were classified independently by two judges as J, NJ, or other, producing subsets 2A and 2B of data set 2. The reliability of the (attenuated) weighted log odds (NJ/J citations) measure was 0.76, a modest improvement over Study 1.

Results

Figure 1 includes Study 2's data. As in Study 1, the (attenuated weight) log odds measure was significantly greater for NJ than for J citing authors, $Fs(1,303) = 12.18$, $p = 0.0006$, and 16.70, $p = 0.0001$, respectively, for subsets 2A and 2B. The odds measure was, averaged over the two subsets, 40.0 per cent higher for NJ than for J citing authors. Again, as in Study 1, the data clearly displayed the pattern that indicates ethnically based citation discrimination.

Discussion

As is true for any hypothesis tests that use nonexperimental data, and as was the case for Study 1, possible alternative interpretations of Study 2's findings abound.

The possibility of a differential assortment artifact was previously noted in discussing the findings of Study 1. To the extent that differential assortment occurs, an artifactual citation bias should appear in any data set in which there is substantial heterogeneity of disciplinary identification. In Study 2, an attempt was made to avoid, or greatly attenuate such artifact, by confining the study to a relatively narrowly defined area of literature. Within that narrow boundary, the possibility of differential assortment should have been greatly reduced relative to the more heterogeneous domain sampled in Study 1[5] Despite this difference between the studies, citation discrimination was observed at virtually the same level in Study 2 as in Study 1.

A second strategy to reduce any differential assortment artifact in Study 1 made use of available information about citing authors' discipline identifications. In Study 2, a greater variety of such supplementary data was obtained, including citing authors' disciplinary affiliations and geographic regions, as well as characteristics of their institutional affiliations (public versus private; religious affiliation versus none; size of enrollment; and population of city or town in which located). These additional variables were used, in additional regression analyses, as dummy-coded predictors of citation log odds ratios. To the extent that citation biases are artifacts of differential assortment, these added predictors should account for significant variance and, consequently, the variance accounted for by citing authors' name ethnicity should be reduced. However, in the several such analyses that were done, the added predictors accounted for little variance, and regression coefficients for the author ethnicity predictor were diminished no more than slightly.

Various factors other than ethnic prejudice might lead researchers to be acquainted selectively with colleagues with whom they share ethnicity. If scientists selectively cite those with whom they have personal acquaintance, such ethnically-based acquaintance patterns could produce the appearance of citation discrimination. A basis for concern about this possibility of *acquaintanceship artifact* was confirmed by examining contingencies between first-author and coauthor ethnicities. When the first author was NJ, the ratio of NJ to J coauthors was 4.53, compared to a ratio of 1.56 when the first author was J[6]. Two tests were conducted in an attempt to evaluate the contribution of possible acquaintanceship artifact to observed citation discrimination.

The first test used the total number of citations in an article as a moderating variable. It was assumed that, the fewer the citations in an article, the more likely it was that any single citation was of a personal acquaintance. This assumption yields the expectation of a negative relation between number of citations and the discrimination index. That is, the more citations in an article, the less the article should display acquaintance-mediated ethnic citation bias. To the contrary, however, the observed relation between number of citations and the discrimination index was generally nonsignificant, and was weakly positive in four of the five data subsets

[5] It may be suggested that NJ and J prejudice researchers could be working on different topics within the domain of prejudice — for example, the J researchers might more likely be working on antisemitism. There is, however, little support for this speculation derivable from examination of article titles. For both NJ and J authors, the major research topic for articles in data set 2 was racial prejudice.

[6] After this association was discovered, citations by article authors of their coauthors were examined to see if they could contribute to the observed level of ethnic discrimination in citations. However, citations of coauthors were much too infrequent to produce more than a tiny contribution to Figure 1's evidence for ethnically-based citation discrimination. The proportion of all citations that were works by coauthors was only 0.7 per cent.

of the two studies — opposite to the relationship suspected on the basis of acquaintanceship artifact.

The second test for acquaintanceship artifact was based on assuming that, the older the publication date of a cited work, the smaller should be the possibility that the citation is based on personal acquaintance between citing and cited author. In the extreme, other authors who had died before the citing author's career started could not be personal acquaintances. Accordingly, detailed analyses were conducted for two of the data subsets (all of the citations in subset 1A and all of those from articles with more than 60 citations in subset 2A), considering separately citations from the 1980s, 1970s, 1960s, and pre-1960. Because the analysis of these data lacked sufficient power for significance tests, the data were simply combined into two observations for each of the four time periods, one for all J authors and the other for all NJ authors in both subsets. These data are presented in Figure 2, showing self-citations for comparison.

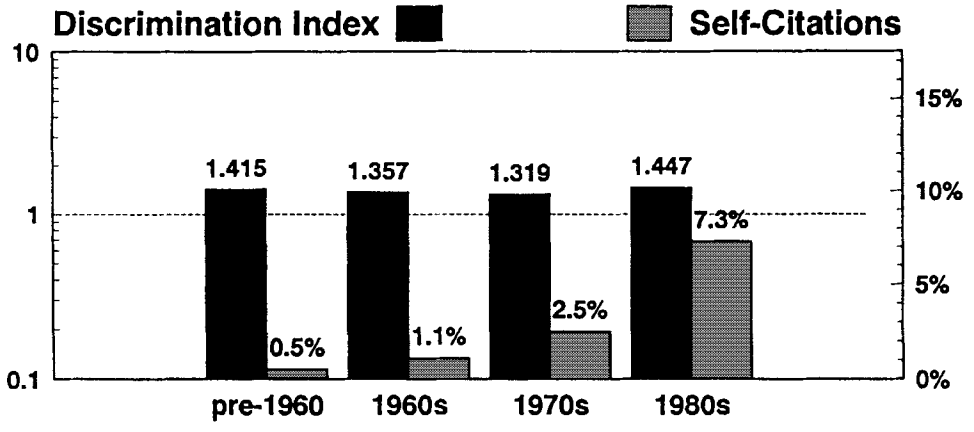

Figure 2. Citation discrimination as a function of publication date of cited articles. Pooled results of subsets 1A and 2A. The stability of the discrimination index across periods, as shown in this plot, makes it implausible that observed citation discrimination was due to selective citing of personal acquaintances. The figures for the four time periods are based, respectively (oldest first), on 2039, 2635, 6390, and 4261 citations

As can be seen in Figure 2, name-ethnicity-based discrimination was apparent at all periods, and was stronger for pre-1960 than for two of the three more recent periods. In contrast, self-citations showed the expected steady decline with age of cited works, and were almost nonexistent for pre-1960. Although these data do not rule out the possibility that acquaintance factors contributed to observed citation discrimination, they are decidedly inconsistent with the supposition that an acquaintanceship artifact can explain the level of citation discrimination shown in Figure 1.

GENERAL DISCUSSION

The analyses of Studies 1 and 2 consistently revealed evidence for ethnically based citation discrimination. In examining citations by nonJewish-named and Jewish-named authors, it was found that each group cited relatively more others whose names fell into the same ethnic category. This citation discrimination could indicate an attitudinal bias that leads to relatively greater citation of own ethnicity However, because the studies used nonexperimental data, there were some plausible alternative interpretations. Two especially plausible alternatives, based respectively on the likely occurrence of *differential assortment* of researchers by ethnicity into research disciplines, and of ethnic selectivity in *acquaintanceship* patterns, were considered in designing the studies and in conducting analyses.

Evaluation of alternative (artifact) interpretations

Evidence concerning the differential assortment interpretation was clearest in Study 2, which started with a narrowly defined topic boundary within which to obtain citations, and showed that (a) within that narrow boundary, citation discrimination was just as great as that within the much broader topic boundary of Study 1, and (b) further, citation discrimination was undiminished within the still more narrowly defined discipline and demographic subregions that were identified in Study 2's data. If differential assortment had contributed to observed citation discrimination, these boundary narrowing strategies should have yielded noticeable reductions in observed discrimination — but they did not.

Study 2 also provided a test of the acquaintanceship interpretation. The strategy was similar to that for differential assortment — identifying portions of the data within which citation discrimination should have been reduced if acquaintanceship was mediating observed discrimination. In particular, for authors publishing in the early 1980s, there should be decreasing personal acquaintance with authors of other studies as one limits attention successively to studies that were published in the 1980s, 1970s, 1960s, and pre-1960. Therefore, acquaintanceship-based citation discrimination should show an orderly reduction as the focus of analysis is restricted successively to cited articles published in these time periods. However, the analysis shown in Figure 2 revealed no reduction in observed discrimination across these periods — ethnic citation discrimination was at about the same level in citations of pre-1960 publications as it was in citations of post-1980 publications.

Who is biased?

In conducting this research, the present authors sought evidence for unconscious operation of ethnic biases in citations, and attempted to do so in a way that would avoid pointing an accusatory finger more at one than the other of the two categories (J and NJ) being studied. The authors have therefore been relieved to discover that, at least within the present data sets, it has not been possible to identify a level of citation odds that serves as a bias-free level, to which the observed values for J and NJ authors might be compared. Consequently, the present research is properly interpreted by observing that its findings may indicate bias by J authors, or by NJ authors, or by both.

Theoretical interpretation

The findings reported here confirm the somewhat distressing expectations that follow from the theoretical analysis of implicit operation of attitudes (Greenwald, 1990; Greenwald and Banaji, in press). In terms of that analysis, (a) evaluative judgments of novel social objects can be influenced by attitudes toward evaluation-irrelevant attribute of those objects, and (b) this biasing can occur without the judge's awareness of the influence.

In the domain of citations, the implicit attitude analysis is applied by treating a newly encountered published work as the 'novel social object'. Evaluation of this social object, the citation, is assumed to be influenceable by attitudes toward evaluation-irrelevant features of the citation, such as the ethnicity of its author's name. Although not invariably so, it is often the case that citing a published work indicates the citer's favourable evaluation of that work. Therefore, the finding that ethnic similarity of citing and cited authors' names is associated with increased likelihood of citing suggests the operation of an attitude that favours the citer's own ethnicity. In interpreting that finding in the present data, the present authors have suggested that citing authors may be unaware not only of the influence of such attitudes on their citation behaviour, but also of the very existence of the ethnic attitudes that appear to be implicitly expressed in citations. This interpretation in terms of unconscious prejudice seems especially justified for Study 2, which attempted to confine its focus to researchers who, by choosing to do research on prejudice, show that they are likely to have an explicit goal of reducing prejudice.

Scholarly evaluations are made in many settings. Therefore, the discrimination found in citations might also appear in other scholarly evaluations. Such evidence might be sought, for example, in evaluations of manuscripts submitted for publication or in evaluations of applications for research grants and fellowships. Because of the availability of a public archive of scholarly citations, hypothesis testing was much more feasible for citations than for these other domains. Going beyond the examination of scholarly evaluations, the methods of the present study could be applied in the settings of many other archival records that include both (a) some indicator of social evaluation, and (b) the names of evaluator and evaluatee in a form that permits ethnic or racial or gender (or other social) categorization of these two parties.

The remainder of the Discussion considers some technical questions that arose in obtaining or interpreting the present findings.

Power of the selective sampling design

A question can be raised about the comparison of the weakly positive (and statistically nonsignificant) effect of ethnicity in the initial pilot study and the very clearly statistically significant effects of the analyses of all data subsets in the two larger studies. This difference can be understood in terms of the different designs of the studies. The effect size ($r = 0.065$) of the pilot study, if used as the estimate of the population effect size, calls for a sample of $n = 1879$ independent citations in order to provide a power of 0.80 (i.e., 80 per cent probability of rejecting the null hypothesis) at $\alpha = 0.05$, two-tailed. In other words, with the pilot study's observed effect size, even a sample more than triple the size of the plot study's sample ($n = 610$) would

have had a substantial chance (20 per cent) of failing to produce a statistically significant result. The design change of selecting citing authors whose names were classifiable unambiguously as J or NJ had the effect of increasing the estimated effect size to $r = 0.214$, which is the r for the subsample ($n = 68$) of the pilot study that included only authors whose names were unambiguously J or NJ. The selective sampling design, by itself, reduces the n required for power of 0.80 by more than a factor of 10 (to 173). The actual numbers of citing authors sampled in Studies 1 and 2 were larger than 173 ($n = 477$ for Study 1 and $n = 356$ for Study 2). Power was further increased by using as many citations as were available for each of these citing authors. *The statistically significant findings of Studies 1 and 2 are therefore entirely consistent with the observed effect that was nonsignificant in the pilot study.* That is, the designs of Studies 1 and 2 provided more than ample power for an effect of the size observed in the pilot study to achieve statistical significance. In confirmation that the sizes of effects in Studies 1 and 2 were as expected on the basis of the pilot study, the observed effect sizes for the five subsets ranged from 0.13 to 0.23, as assessed by the beta for author ethnicity in unweighted regressions analyses. These effect sizes, although in the small to moderate range, are large enough to be of practical significance in a large population of events. As one comparison, for example, they are larger than the effect size that defines the very profitable house advantage on a roulette table.

Possibility of undetected artifact

As already mentioned, two plausible artifactual causes of ethnic discrimination in citations, (a) differential assortment of scientists by ethnicity to research topics, and (b) selective acquaintanceship of researchers based on ethnicity, were considered. Analyses that sought evidence for these alternative interpretations uniformly failed to show that they could explain the observed levels of ethnic discrimination in citations. This lack of evidence for plausible artifacts, together with the consistent observation of ethnic discrimination in all five data subsets of Studies 1 and 2, leaves the hypothesis of attitudinal bias, operating at the level of individual authors, as a remaining plausible explanation. How, it should be asked, can one go further in designing citation studies that surpass the present ones in reducing the possible contributions of artifacts?

In order to minimize the possibility of both differential assortment and acquaintanceship artifact, one might investigate citation pattern differences within ethnically heterogeneous pairs of researchers who are matched for research domain (thereby avoiding any differential assortment problem) and for date and institution of PhD (making it likely that they have the similar groups of professional acquaintances). Or, one might identify ethnically heterogeneous pairs of researchers who have published collaboratively, but with different orders of authorship, so that one could examine whether the choice of which of the pair to cite as first author is influenced by ethnicity of citing author. Unfortunately, such studies are quite difficult to do, both because there is no readily available archive that identifies these special types of researcher pairs, and because of the need for a large number of these special pairs in order to provide adequate statistical power. More troublesome, even if these pairs could be easily identified, studies of them would still remain open to the possibility of acquaintanceship artifact — the members of each pair would certainly have

different sets of acquaintances, and these sets of acquaintances might well be ethnically different.

Consideration of still other possible designs leaves the conclusion that *further citation studies — no matter how carefully designed — almost certainly will not remove doubts that remain regarding interpretation of the present studies.* One might ordinarily insert here a call for experimental studies. In the present case, however, this call is unnecessary, because there have already been several effective experimental demonstrations that make plausible the unconscious operation of prejudice (reviewed by Banaji and Greenwald, 1993b; Crosby, Bromley and Saxe 1980; Gaertner and Dovidio, 1986). Also, development of new methods for indirect measurement of stereotypes and prejudices in experimental studies continues actively (see Banaji and Greenwald, 1993a; Dovidio and Fazio, 1992). The present investigation, indeed, sought to determine whether unconscious prejudice of a type that has been revealed repeatedly in experimental studies was also identifiable in an interesting domain outside the laboratory. Perhaps the most suitable follow-up on the present observations is to seek evidence for similar biases in other nonlaboratory domains.

Citation discrimination and estimates of research productivity

The finding of ethnic discrimination in citations indicates that citation-based measures may underrepresent productivity of some minorities. *Even assuming that discrimination in favour of own ethnicity is engaged in equally by minority and majority groups, citation-based measures will underrepresent minority productivity*, because of the greater volume of citations by the majority Note that this concern about the possibility of error in using citations as comparative productivity measures is valid even if observed citation discrimination is assumed to be explained by an attitudinally benign acquaintanceship artifact, rather than by unconscious prejudicial attitude.

Previous studies have often reported that members of a social category may self-direct the same discrimination that is exercised against it by members of another category, as when women discriminate against women (Goldberg, 1968; Swim, Borgida, Maruyama and Myers, 1989), or blacks against blacks (Clark and Clark, 1947). If such a phenomenon occurs also in the case of citations (for example, if women undercite women), then citation rates might severely underestimate the scientific productivity of members of some groups.

Conclusion

The present findings add to previous research that has demonstrated discrimination by persons who qualify as nonprejudiced on the basis of apparently honest self-reports (see reviews in Crosby *et al.*, 1980; Devine, 1989; Gaertner and Dovidio, 1986). Interpretations of some of the earliest such findings (e.g. DeFleur and Westie, 1958) focused on problems of measurement and validity for the attitude construct that provides the basis for many definitions of prejudice. The interpretations of more recent findings, including the present ones, emphasize instead the extent to which discrimination may be under the control of unconscious or implicit attitudes, rather than conscious attitudes.

ACKNOWLEDGMENTS

The authors thank Katharina Engnell and Alice Barnes for assisting in data analysis, and Mahzarin R. Banaji, Patricia G. Devine, Samuel L. Gaertner, Richard Gonzalez, Mark R. Klinger, Elliott Lowy, Thomas F Pettigrew, and three anonymous reviewers for comments on earlier drafts. This research was partly supported by grants from the National Science Foundation (DBC-9205890) and the National Institute of Mental Health (MH41328).

REFERENCES

Agresti, A. (1990). *Categorical Data Analysis*, Wiley, New York.

Banaji, M. R. and Greenwald, A. G. (1993a). 'Implicit gender stereotyping in judgments of fame' Unpublished manuscript, Yale University.

Banaji, M. R. and Greenwald, A. G. (1993b). 'Implicit stereotyping and prejudice' In: Zanna, M. P and Olson, J. M. (Eds) *The Psychology of Prejudice, The Ontario Symposium*, Vol. 7, Erlbaum, Hillsdale, NJ, pp. 55–76.

Baxter, G. W (1973). 'Prejudiced liberals? Race and information effects in a two-person game', *Journal of Conflict Resolution*, **17**: 131–161.

Chatterjee, S. and Price, B. (1977). *Regression Analysis by Example*, Wiley, New York.

Clark, K. B. and Clark, M. P (1947). 'Racial identification and preference in Negro children' In: Newcomb, T M. and Hartley, E. L. (Eds) *Readings in Social Psychology*, Holt, New York, pp. 169–178.

Cooper, W H. (1981). 'Ubiquitous halo' *Psychological Bulletin*, **90**: 218–244.

Crosby, F., Bromley, S. and Saxe, L. (1980). 'Recent unobtrusive studies of black and white discrimination and prejudice: A literature review', *Psychological Bulletin*, **87**: 546–563.

DeFleur, M. L. and Westie, F R. (1958). 'Verbal attitudes and overt acts: An experiment on the salience of attitudes', *Social Forces*, **27**: 667–673.

Devine, P G. (1989). 'Stereotypes and prejudice: Their automatic and controlled components', *Journal of Personality and Social Psychology*, **56**: 5–18.

Dovidio, J. F and Fazio, R. H. (1992). 'New technologies for the direct and indirect assessment of attitudes' In: Tanur, J. M. (Ed.) *Questions about Survey Questions: Meaning, Memory, Expression, and Social Interactions in Surveys*, Russell Sage Foundation, New York, pp. 204–237

Eagly, A. H., Ashmore, R. D., Makhijani, M. G. and Longo, L. C. (1991). 'What is beautiful is good, but .. A meta-analytic review of research on the physical attractiveness stereotype', *Psychological Bulletin*, **110**: 109–128.

Gaertner, S. L. (1976). 'Nonreactive measures in racial attitude research: a focus on "liberals"' In Katz, P A. (Ed.) *Toward the Elimination of Racism*, Pergamon Press, New York, pp. 183–211.

Gaertner, S. L. and Dovidio, J. F (1986). 'The aversive form of racism' In: Dovidio, J. F and Gaertner, S. L. (Eds) *Prejudice, Racism, and Discrimination*, Academic Press, Orlando, FL, pp. 61–89

Goldberg, P (1968, April). 'Are women prejudiced against women?' *Transaction*, **5**: 28–30.

Greenwald, A. G. (1990). 'What cognitive representations underlie social attitudes?' *Bulletin of the Psychonomic Society*, **28**: 254–260.

Greenwald, A. G. and Banaji, M.R. (in press). 'Implicit social cognition: Attitudes, self-esteem, and sterotypes' *Psychological Review*.

Landy, D. and Sigall, H. (1974). 'Beauty is talent: Task evaluation as a function of the performer's physical attractiveness' *Journal of Personality and Social Psychology*, **32**: 311–329

Linn, L. S. (1965). 'Verbal attitudes and overt behavior: A study of racial discrimination', *Social Forces*, **43**: 353–364.

Swim, J., Borgida, E., Maruyama, G. and Myers, D. G. (1989). 'Joan McKay versus John McKay: Do gender stereotypes bias evaluations?' *Psychological Bulletin*, **105**: 409–429

APPENDIX — TECHNICAL NOTES ON DATA ANALYSES

Weighting of cases

The single citing author is the appropriate unit of analysis for the citation odds measure [(number of citations of NJ authors) − (number of citations of J authors)] that was used in the present data analyses. The total number of citations (NJ + J) on which the ratio was based ranged from 1 to approximately 200 (see Figure 3). This caused a problem: Should cases (i.e. authors) with few citations be relied on as strongly as those with more citations in estimating effects of author ethnicity on the citation odds measure?

The simplest way to analyse the data is to conduct an *unweighted* analysis (the ordinary form of regression analysis), in which the data for each case are treated equally, regardless of the number of citations in the case. A problem with this strategy is that power is compromised by the relatively high variability of the log odds index for cases that are based on small numbers of citations. This problem is solvable with a weighted analysis, a standard method for which is to give each case a weight approximately proportional to the number of citations it contains. For this *standard weight* analysis the weight, w_i, for each case is $n_i p_i\ (1 - p_i)$, where n_i is the total number of J and NJ citations for the ith case, p_i is the proportion of NJ citations in that case, and $(1 - p_i)$ is the case's proportion of J citations (Chatterjee and Price, 1977). The criterion variable for each case in this standard weight analysis is the log odds index multiplied by the case weight.

A problem with the weighted analysis is that the highly skewed distribution of numbers of citations (see Figure 3) allows the few cases that have largest numbers of citations to dominate the analysis. In order to cope with this problem another, intermediate, alternative was considered — an *attenuated weight* analysis that reduced the relative weights of larger cases. For this analysis, the weight assigned to each case was computed by replacing n_i with its square root and, in addition, using $n_i = 50$ for all cases with $n_i > 50$.

The analysis given in the main text used the third (attenuated weight) strategy Tests of significance were conducted in the standard format for weighted regression. In particular, the criterion was the log odds measure multiplied by the weight, the regression was computed 'through the origin' (no constant term is included), and case weight was entered as a predictor prior to entry of the theoretically focal predictor of citing author's ethnicity (dummy coded as 0 or 1). For analyses that included additional predictors (demographic categories and/or disciplinary affiliation), the additional predictors were dummy coded and entered prior to citing author's ethnicity, which was always the last-entered predictor.

Agreement of results from multiple analyses

In addition to the analysis (described in the main text) of the attenuated weight log odds index, analyses of all data subsets were also conducted using both the unweighted and the standard weight log odds indexes. With one exception over all data subsets, the three analysis strategies yielded the same conclusions (i.e. the same patterns of significant and nonsignificant findings, using two-tailed $\alpha = 0.05$ as the significance criterion). The exception was that the standard weight analysis

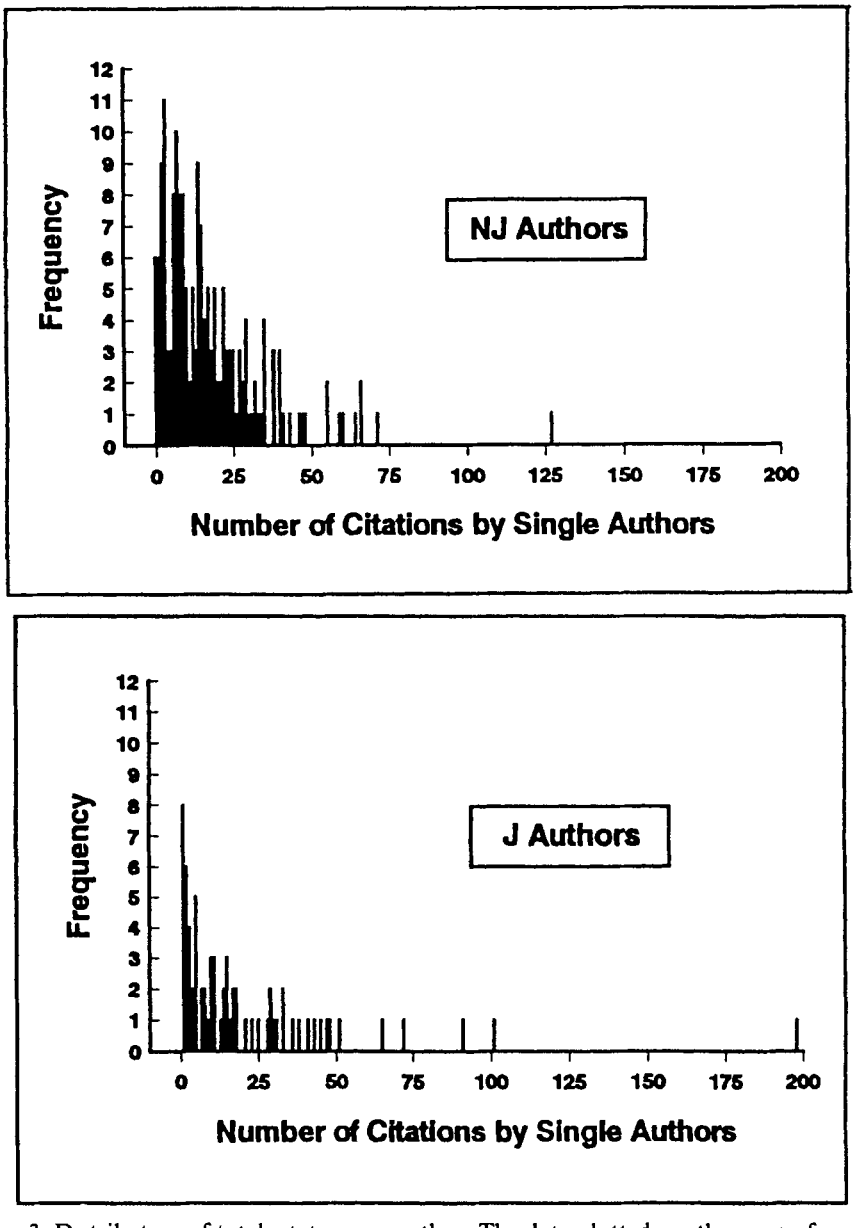

Figure 3. Distribution of total citation per author. The data plotted are the sums of numbers of citations categorized as either J or NJ in subset 1C. Similarly highly skewed distributions characterized the other subsets examined in this article

for subset 1C yielded a nonsignificant effect of ethnicity on the log odds index, whereas the other two analyses both yielded a significant ethnicity effect. Closer examination of the data indicated that this discrepancy was a consequence of the single largest case in each ethnic category deviating from the general pattern of the remaining cases. (These were the two largest cases shown in Figure 3.) Because

of their large size, these two cases amounted to 7 per cent of the weighted sample, while being less than 1 per cent of the unweighted sample.

Another assessment of agreement among the three analyses was obtained by examining the effect of author ethnicity on the discrimination index, averaged over the data subsets within each study (The discrimination index is the mean value of the citation odds index for NJ researchers, divided by the mean for J researchers. Values of this index greater than 1.0 indicate greater citation of own ethnicity) For Study 1, the average value of the discrimination index (across three subsets) for the attenuated weight analysis was 1.408 (indicating 40.8 per cent more citing of either J or NJ authors by members of the same group than by members of the other group — this figure was reported in the main text). For comparison, the unweighted analysis gave an average index of 1.455 and the standard weight analysis gave an average index of 1.406. For Study 2, the attenuated weight analysis yielded an average index (over the two subsets) of 1.400. For comparison, the unweighted analysis gave an average index of 1.613 and the standard weight analysis yielded an average index of 1.436. In sum, there was generally good agreement among the analyses, and the attenuated weight analysis strategy tended to yield slightly lower estimates of discrimination than did the other two.

Treatment of cases with zero J or NJ citations

The major dependent variable was the logarithm of a ratio. A zero in the denominator makes the ratio uncomputable, and a zero in the numerator makes the logarithm uncomputable. Accordingly, when a case had either zero NJ citations or zero J citations, the dependent variable was uncomputable. Standard fix-up solutions for this problem are either to add 0.5 to all numerators and denominators, or to replace zeros with a very small constant, such as 10^{-8} (Agresti, 1990). The strategy used for the analyses reported in the main text was intermediate between these two — adding a constant to the numerators and denominators of all cases for which one of these was zero. The constants added to numerators and denominators were, respectively, the overall proportions of citations of NJ and J authors in the data subset being analysed. In effect, this procedure adds a single citation to cases containing a zero numerator or denominator, allocating a fraction of that citation (equal to the overall fraction of NJ citations in the data subset) to the numerator, and similarly for the denominator (J citations). This strategy is conservative, in the sense that it reduces the mean difference on the criterion between the subsamples of NJ and J citing authors. At the same time, this strategy produces less distortion in aggregate indexes and in statistical test outcomes than do either the add-0.5-to-all-numerators-and-denominators strategy or the replace-zero-with-a-tiny-constant strategy (as demonstrated in simulations described by Banaji and Greenwald, 1993a). The expected conservatism of the adopted fix-up procedure was confirmed by observing that it produced numerically smaller values of test statistics than did either of the other two procedures.