

Academy Of Management Journal

EDITOR

R. Duane Ireland
Texas A&M University
direland@mays.tamu.edu

ASSOCIATE EDITORS

Peter A. Bamberger
Technion - Israel Institute of Technology
peterb@tx.technion.ac.il

Jason A. Colquitt
University of Florida
jason.colquitt@cba.ufl.edu

K. Michele Kacmar
University of Alabama
amjmk@ua.edu

David J. Ketchen, Jr.
Auburn University
ketchda@auburn.edu

Elizabeth W. Morrison
New York University
emorriso@stern.nyu.edu

Michael G. Pratt
University of Illinois at Urbana-Champaign
amjpratt@business.uiuc.edu

Wm. Gerard Sanders
Brigham Young University
gsanders@byu.edu

Wenpin Tsai
The Pennsylvania State University
wtsai@psu.edu

MANAGING EDITOR
Michael P. Malgrande
Academy of Management
mmalgrande@pace.edu
914-923-2607
Fax: 914-923-2615

PRODUCTION EDITOR
COPY EDITOR
Persephone Doliner
Ithaca, NY
pdoliner@twcny.rr.com
607-277-5283

Sept. 6, 2008

Dear Professor Hekman:

On behalf of the entire editorial team, I want to thank you for submitting your paper to AMJ for publication consideration. Titled, "An examination of whether and how racial and gender biases influence customer satisfaction ratings," your paper is logged into our system as AMJ-2008-0445. Please refer to this number in all future communications associated with this particular paper.

Your paper was sent to three scholars who are intimately familiar with the area of inquiry with which your research questions are concerned. The three scholars have completed their reviews. For your convenience, I am attaching the reviewers' commentaries to the end of this decision letter. All three reviewers provide you with excellent observations as well as with a very constructive set of recommendations.

As you will see, the general response from all three reviewers to your work is encouraging. In the first paragraph of feedback, Reviewer 1 highlights several strengths of the study, "including the three study design that incorporates both field and lab research, the use of objective and subjective indicators of performance, and the variety of organizational settings." Reviewer 2 (overview) cites many of the same strengths as well as several additional ones, noting that "the samples and data were quite impressive (especially for studies 1 and 3), the results were consistent and clear, and the paper was generally well written." Finally Reviewer 3 (overview) again citing these same assets, adds that the paper's integration of field and experimental methods provides a "deeper understanding of the researched phenomenon."

However, please observe that, despite our enthusiasm for your topic, the reviewers and I feel that there are quite a number of significant theory-based and methodological problems that, in combination, have the potential to limit the contribution of this study to the

literature. While all three reviewers raise questions concerning inconsistencies or ambiguities in one or more of the studies conducted, the most significant concerns are theoretical. In particular, while the potential for an empirical contribution here is clear, the reviewers and I are skeptical about the theoretical contribution. Additionally, the reviewers (particularly R1) and I are not convinced that the current analyses map onto the theory and hypotheses. As a result, we are presented with findings (albeit, very interesting ones) which lack a strong theoretical basis. So, while as suggested above, the reviewers and I find the research question interesting, the dataset impressive, and the findings intriguing, some very basic questions need to be answered before we can accurately assess this study's theoretical and empirical contribution.

I will summarize the most critical of our concerns in my detailed feedback below. However, at this point let me say that given the potential offered by your line of inquiry and unique dataset, I am prepared to offer you the opportunity to prepare a revision of your work that can be submitted to *AMJ* for a second-round review. Please consider this opportunity very carefully as a successful revision is likely to require a great deal of work (including, for example, a more detailed specification of the theory to address the mechanisms underlying the expected effects, the re-specification of the hypotheses and accordingly, the re-running of the analysis, the re-doing of the second study, and even, perhaps, the collection and/or integration of additional data from the field). In this regard, while this manuscript may very well offer significant potential, there still remains a **VERY HIGH DEGREE OF RISK**. I am saying this merely in the spirit of forthrightness; indeed, I have no desire whatsoever to discourage you from pursuing the revision along the lines that is communicated below in this letter and along the lines of the actions suggested by the reviewers' recommendations.

Before proceeding to detail the concerns that I am asking you to address through the revision, let me first provide information regarding the procedures we are asking you to follow with regard to the handling of the revision process.

PROCEDURES FOR REVISING YOUR MANUSCRIPT

In revising your manuscript, please carefully consider each reviewer comment and pay particular attention to the points mentioned below in this decision letter. For the past two years, *AMJ*'s policy has requested that revision efforts and "responses to reviewers and the action editor" documents concentrate on points covered in the decision letter. While this approach has merits, we (the current *AMJ* editorial team) are modifying it to ask that you deal with *all* issues raised by the reviewers and the action editor while revising your manuscript and that you provide point-by-point responses to explain how you have done so. We believe that having you explicitly respond to all the issues raised by the reviewers and action editor puts you in the best possible position to achieve a favorable outcome. Moreover, we believe that providing point-by-point responses allows you to precisely explain each action you have taken to deal with issues the reviewers and action editor brought to your attention. However, *concisely* explaining the actions you have taken is desirable in that such explanations save reviewers' time while ensuring that your

responses highlight the actions you have taken to deal with their concerns. The type of responses we are requesting from you means that extended discussions of tangential issues should be avoided as should reproductions of large blocks of text from the paper within the responses document. While not imposing a page limit for the responses document, we ask that you be as concise as possible as you develop your point-by-point responses. To this end, if the same point is raised by the action editor and/or one or more reviewers, you should provide a detailed response only once and then refer the other readers (i.e., the action editor and/or the reviewers) to the initial response you provided regarding a particular (and commonly-shared) issue or concern. May I also ask that you copy each critique in italics into this document and place your response directly beneath it? This will make it easier for the reviewers and me to know what critique it is that you are addressing without having to flip back to our original comments. The responses document should appear at the end of the revised manuscript, beginning on a separate page.

Instructions for Resubmission on Manuscript Central

Once your revised manuscript is completed and you have tagged your Executive Summary (i.e., the letter indicating how you responded to our concerns) to its end, log into <http://mc.manuscriptcentral.com/amj> and enter your Author Center, where you will find your manuscript titled listed under “Manuscripts with Decisions.” Under “Actions,” click on “Create a Revision.” Your manuscript number will have been modified to denote a revision. **IMPORTANT:** Your original files are available to you when you upload your revised paper. Please delete any redundant files before completing the submission.

Our desire is to facilitate timely publication of papers accepted for inclusion in *AMJ*. As such, please upload your revision on the Manuscript Central system within four months. Please let me know if you need more than four months to complete the revision process. We will work with you regarding the time needed to complete a revision while maintaining *AMJ*'s commitment to timely editorial processes and decisions. Similarly, please let me know as soon as possible if you opt not to revise this manuscript.

SPECIFIC CONCERNS THAT NEED TO BE ADDRESSED

With this information expressed to you, let's now turn our attention to providing feedback to you regarding the reactions the reviewers and I have about this version of your work. I will present a numbered set of points to which you should respond when preparing the revision:

1. **THEORETICAL CONTRIBUTION:** As is noted in *AMJ*'s information for contributors (http://journals.aonline.org/amj/contributor_information.html), to be published in *AMJ*, “a manuscript must make strong empirical and theoretical contributions and highlight the significance of those contributions to the management field.” That in mind, a very significant weakness of the current manuscript is its limited theoretical contribution. This is evident in both the introduction (R3, point

2), and the discussion (R1, point 16, R2, point 1; R3, point 13). In terms of the introduction, R3 (point 2) writes, "I think an important challenge of the current paper is to further think how the important research questions raised can be framed theoretically and how they can contribute not only to empirical knowledge, but also further our theory in this field." One option, as suggested by this reviewer, is to develop a broader theory focusing on the role that customers have in shaping organizational processes. However, I recognize that such a theme may go beyond the data in hand. Another option might be to focus on the mechanisms driving the expected effects. The mechanisms currently laid out on page 7 are, as you suggest, rather obvious ("as any human resource management textbook explains...") and not well developed. For example, Reviewer 1 (point 4) wonders whether it is really accountability or anonymity that may be driving the bias. I encourage you to "dig deeper" in exploring the nature of these mechanisms and some of the "levers" that may cue them. For example, in study three, might homophily theory help in understanding how the demographic compositions of a workforce might influence social integration and staff cohesion, resulting in some palpable difference in the quality of service? To the degree that you can identify cues having an empirical referent for which you have data, I would encourage you to generate and test one or two contingency hypotheses, exploring when such spillover effects may be more powerful and when they may be weaker. Aside from offering you a direct response to the reviewer critiques (i.e., R2, point 4a; R3, points 3 and 4), such an analysis would offer a clear and indisputable theoretical contribution.

In terms of the discussion, as Reviewer 1 (point 16) notes, "additional thought about the constructs and the implications of this research for theory is needed." Clearly the nature of the revised discussion will be shaped by the revised theory presented in the introduction. However, it is clear that the discussion must go beyond a simple restatement of the empirical findings and their practical implications. For example, Reviewer 2 (point 1) asks that you consider expanding on the implications of your findings with respect to the "burgeoning literature on the diversity to performance relationship? And how does your work contribute to relevant literature on customer service organizations and service quality?" Reviewer 3 (point 13) suggests that the discussion may be a good place to discuss the implications of your findings for a nascent theory of implicit customer influences on organizational processes. Finally, as noted by R2 (point 1), "you did not discuss the limitations of your studies (except briefly after study 1 in order to set up the reader for study 2)." Consequently, while the manuscript is already quite lengthy, I would like to see you incorporate a brief review of the limitations (and I believe that the reviewers offer you what to work with in this regard) in the discussion.

2. **THE THEORY ITSELF:** The reviewers identified a number of critical concerns with the theory which you present, and in particular, the degree to which the analyses map on to the hypotheses presented. Reviewers 1 (point 1) and 2 (points 3a and 5d) identify a number of these problems, but I have additional ones. First, the theory and hypotheses suggest the existence of a main effect of employee gender/race (or alternatively, the gender/racial composition of the facility) on customer ratings.

However, particularly in studies 1 and 3, the focus is on the role of employee gender/race in conditioning the link between objective performance and customer ratings. This is fine, but if that's the case, there is the need for a deeper level of theoretical analysis up front to explain why this relationship is conditioned by employee gender/race, and what the nature of this conditioning effect is (i.e., does it reverse the direction of the relationship or simply amplify/attenuate the slope?). Second, related to this point about the conditioning effect of employee gender/race, there is a need to better explain why this conditioning effect is likely to be what you posit it to be. As R1 (point 1) notes, while "research shows that positive ratings for women and minorities are given only when the quality of work is obviously good, ...it makes little sense that individuals would denigrate good care or good service or good facilities just because minorities and women delivered the care/service or work at the facility." I agree with this reviewer and want to emphasize the importance of not only specifying the hypothesis in the form of an interaction (perhaps in addition to a main effect hypothesis if you wish), but clearly grounding this hypothesis on either existent theory or on new theory which is backed by a combination of solid reason and prior empirical research.

Additionally, the reviewers (R1, point 5; R2, points 3a and 5d) wonder why there is no theory regarding the impact of being in a lower status on *both* of the demographic dimensions examined (or alternatively, having a larger proportion of female minorities in the workforce). Should we expect the main or interaction effects of employee demography to be even stronger in such conditions or are supplementary low status attributes unlikely to matter? Obviously, the testing of such a hypothesis is rather straightforward, involving the testing of a 3-way interaction between gender, race and objective performance.

3. **METHOD & RESULTS:** As you will see from the attached reviewer comments, the bulk of the reviewer critiques relate to methodological problems, inconsistencies and ambiguities in each of the three studies. Individually, most of these critiques are, in the scheme of things, of minor to moderate concern. However, taken in combination, they raise my level of uncertainty and doubt with respect to the findings presented. Consequently, care needs to be taken to address each of the critiques raised. As they are too numerous to discuss in detail in this letter, I will only highlight themes as well as those particular critiques that I found most concerning.
 - a. **Suppression Effect** – Relating to Study 1, Reviewer 1 (point 2) makes a convincing argument that your findings may be the result of a suppression effect. To what degree can you assure us that this is not the case?
 - b. **Omitted Variables** – In each study, it is possible that demography is masking the effects of other non-specified variables. For example, consistent with the remarks of Reviewers 1 (point 2) and 3 (points 8 & 9), in Study 1, I wonder whether the conditioning effects of nonwhite and female are simply masking the underlying conditioning effects of tenure? Given the rather strong inverse correlations between these variables and tenure, and recognizing that the objective effects of performance are more likely to be linked with customer ratings to the degree that customers have more experience with the provider, I

would like to see whether the observed effects still hold when controlling for the interaction of tenure with objective performance. Also, for the reasons laid out in his/her critique, in Study 1, Reviewer 2 (point 3c) would like to see you control for customer gender and race. Finally, in Studies 1 and 3, I can see a possible confound between demography/demographic composition and human capital. Consequently:

- i. In Study 1, I would like to see the models re-specified and re-tested to include whether or not the provider completed a fellowship and/or is a specialist (as opposed to general practitioner). My concern here is that while customer ratings may be higher for specialists, non-whites and females may be less likely to work in such positions. Do the interactions retain significance when the interaction between objective performance and provider specialization are taken into account?
 - ii. In Study 3, I would like to see the models re-specified and re-tested to include education, occupational tenure and organizational tenure. My concern here is that facilities with a higher proportion of minorities may be adopting a more “low-end” approach to service provision, with lower customer service ratings being more in response to that approach, than representing bias or discrimination. In addition, since, as you note in the discussion, women and non-whites tend to be underpaid relative to white males, it may be that facilities with a high proportion of non-whites and females are implementing a low-cost/low quality strategy while facilities with those facilities employing more white males are focusing on lower headcounts but high-end services provided by individuals who are either more educated, more experienced or more available (to work overtime). Consequently, given that you already control for headcount, I would like to see some sort of control for labor costs in the form of total facility payroll or perhaps median hourly or weekly pay. Do the interactions stay significant when the interaction of objective performance and these human capital indicators are taken into account?
- c. Inconsistencies and Missing Information – Reviewers 1 (points 6 and 13) and 3 (point 12) point out a number of places where there are inconsistencies between the text and the tables or between one study and another, as well as places where critical information (particularly regarding the nature of the measures) is missing, or the presented information simply makes little sense. Please address each of these points. Additionally, as Reviewer 3 (point 6) notes, there may be an inconsistency between the assumption that participants in experiments are less biased than those in field research, and findings in the literature. I suggest that you either drop this assumption, or rephrase it in order to better account for the findings mentioned.
 - d. Mode of Analysis -- As noted by Reviewer 1 (points 7, 10 and 13), there are questions regarding the mode of the data analyses in all three of the studies presented. For example, its unclear why in studies 1 and 3, the interactions for one demographic parameter were run separately in Model 2, with the interactions for the other parameter run along with the first in Model 3.

Building on the suggestions of Reviewer 1, I suggest first running a control model, then a model including the main effects of the two demographic parameters (assuming you keep a main effect hypothesis), then separate models for the interactions with each demographic parameter, and finally a model testing the 3-way interaction noted earlier.

- e. Priming Effects in Study 2 and other threats to internal validity – Reviewer 2 (points 4b-4d) raises a very significant concern regarding the risk of a priming effect. I agree with this reviewer that it may be worthwhile considering the re-doing of this experiment with the IAT measurement coming *after* participants rate the employees. Additionally, given the nature of the items in the customer satisfaction, like this reviewer, I wonder how much satisfaction is being influenced by the videos and how much it is being influenced by prior experience in the bookstore. Might it be possible to control for the variance in participants' recent patronage of the bookstore and/or their recent positive/negative book store experiences?
- f. Threats to External Validity – Reviewers 1 (point 12) and 2 (points 2 and 5c) raise significant questions of external validity, pointing to an issue that is raised on page 6 of the manuscript, namely that of occupational or role congruence. As noted by Reviewer 1, at least in study 2, this issue might be addressed by demonstrating, based on data collected “from individuals other than your subjects,” that the non-white and female employees in the videos were engaged in work roles that were not incongruent to their demographic profile. As for Study 1, you may want to consider running an analyses comparing those providers in more congruent roles with those in more incongruent roles. However, if this is not feasible, you may have little choice but to simply acknowledge some of the possible limitations to external validity raised by Reviewer 2.
- g. Marginal significance and over-interpretation in Study 3 -- I agree with Reviewer 1 (point 14) that there really is no such thing as “marginal significance.” Given that sample size, statistical significance should be indicated only when $p < 0.05$. I simply see little reason to report results with p values of .05 to .10 as “marginally significant.”

Let me close by again thank you for submitting your work to AMJ and for the opportunity to provide our collective feedback to you. We are pleased about the editorial decision being submitted to you and look forward to receiving the revision we are requesting of you.

All the best,

Peter Bamberger

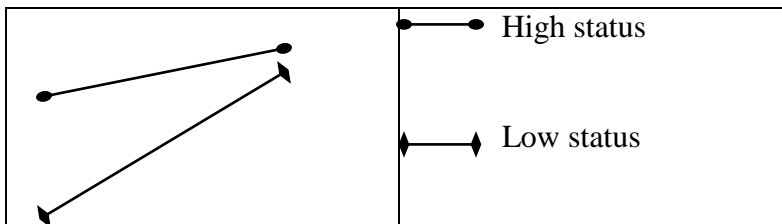
Peter Bamberger
Associate Editor, Academy of Management Journal
peterb@tx.technion.ac.il

R1

This manuscript addresses the questions of whether race and gender biases are related to lower customer satisfaction ratings of employees who are nonwhite or female. There are a number of strengths of this manuscript, including the three study design that incorporates both field and lab research, the use of objective and subjective indicators of performance, and the variety of organizational settings. Although there are some places in the manuscript that could use greater clarity (which I will note below), this paper is also fairly easy to follow and well-written.

However, there are a number of questions that remain about this research.

(1) The form of the interactions that were shown in Studies 1 and 3 do not seem consistent with the theoretical arguments set forth in the paper. As you described in the introduction, research shows that positive ratings for women and minorities are given only when the quality of work is obviously good. Additionally, it makes little sense that individuals would denigrate good care or good service or good facilities just because minorities and women delivered the care/service or work at the facility. Thus, the interactions that you found, though consistent across Studies 1 and 3 do not seem to correspond to what the theoretical interaction should look like. From your introduction, I anticipated a main effect for high status vs. low status, and a main effect for high performance vs. low performance, and an interaction that would have different INCREASING slopes for both high status and low status personnel, as the data move from low quality performance to high quality performance. That is, it might have looked something like this:



(2) Looking at the results in Tables 1 and 2 for Study 1, you seem to have a suppressor effect occurring. The zero order correlations for panel age and physician age with satisfaction are of the opposite sign in the regression; additionally, panel's chronic sickness and physician tenure have much higher regression coefficients than they do zero order correlations. It seems that there is a good chance that these four variables (panel age and chronic sickness, physician age and tenure) are related. The two sets (panel variables as one set, physician variables as the other) are fairly obviously related within set (especially employee age/tenure, a robust finding in all of OB). But also, more tenured physicians might have patients who have been in their panel for some time, and therefore they are older and also more likely to have a chronic illness. The zero order correlations suggest some support for these ideas.

(3) The “introduction” section of Study 1 would be better in the general introduction to the manuscript. It generally does not provide specific insights to the hypothesis addressed in Study 1, but rather adds further information about the general theoretical framework that all three studies address.

(4) On page 10, is it really “lack of accountability” that leads to biased ratings, or is the ability to be anonymous, thus allowing individuals to slip “backstage” (to use Feagin’s term of art)? I am not suggesting that lack of accountability is unimportant, but rather that accountability isn’t an issue (i.e., something that customers know to be aware of, rather than something they choose not to be) and anonymity is common.

(5) Hypothesis 1 does not account for female minorities as experiencing a double jeopardy. Your introduction skirts the issue, but the theoretical work would probably be better developed if you addressed this issue. Additionally, if possible, three-way interactions (sex, race, and objective performance measures) might be illuminating in Studies 1 and 3.

(6) Table 1 and the methods section for Study 1 have a few inconsistencies, as well as some needs for more clarity. First, white/nonwhite and male/female cannot be coded 0/1 if the means (Table 1) are above 1.0. Second, the physician quality variable has a mean of zero in Table 1, yet the description in the methods section refers to a prescription rate. If you standardized, that’s acceptable, but should be reported. Finally, greater clarity is needed regarding (a) whether it was possible to determine if the emails sent by the physicians were patient/business emails or personal emails and (b) what exactly “average time until each physician’s third available appointment for the quarter” means. This latter question is partly a wording issue. Is it that, after the quarter ended, records were scoured to determine when the third “open” (unassigned and never assigned to an appointment, i.e., ignoring last minute cancellations) appointment was, across all days? Or averaged on each day? And was this controlled by the number of hours into the work schedule that this slot appeared? And why is the third appointment the appropriate metric?

(7) In Study 1, the two-way interactions involving sex were entered first, and then the two-way interactions involving race. IN Study 3, the opposite was done. Why? It seems that you would want to do each (sex interactions, race interactions) as a second step, and have both in as the third step, so you can determine the incremental validity of each over the other.

(8) In the regression tables, you should also report adjusted R-squared, as several of the regression models (esp. in Study 1) have quite a few variables.

(9) On page 19, you wrote that “Biases against African-Americans may be more or less negative than those associated with Asians...” It seems that whether it is more or less makes a difference to your arguments, and it also seems that the research literature has a fairly clear answer on which direction that relationship is.

(10) It's not clear why Study 2 does not use a 2x2 (sex X race) factorial design. Further, your description of dummy codes in the methods section is not very clear (p. 24). Are you just saying (as it seems in Table 3) that you looked at white male-white female target in one regression, and white male-black male targets in another, separate regression? If not, then you might have a problem with your analysis, as the description seems to suggest that one dummy code is 0=black male, 1=white male or white female, whereas the other dummy code is 0= white female, 1=white male or black male. This would not allow you to separate the groups that you want to compare. It might be easier (and result in the same results) if you considered using the ANCOVA function in your statistical package, instead of the regression function.

(11) Throughout Study 2, there is some confusion on whether or not you are measuring ratings of the target employees. On page 21, you suggest that you aren't. On page 23 at the top, you state that you are. Then on page 24, it turns out that the employee ratings are manipulation checks (even though many of the items seem to read like a satisfaction scale as much as a performance scale). I don't disagree with your arguments about why people might show their biases toward the store rather than the people, but at the same time this section whiffs of post-hoc justification of not treating employee ratings as a DV. That might be unfair of me to state, and I am certainly not accusing you of hypothesizing post-results. Instead, I am suggesting that the way that this information is presented makes it seem this way. Greater attention to why that manipulation check is actually a manipulation check (and not a satisfaction measure) would help as well (i.e., tie specific items back to specific behaviors in the videos).

(11) It might be useful to acknowledge the debate about the usefulness and meaning of IATs.

(12) On page 27, top, an additional reason why women received worse ratings could also be that for this job, women are more role incongruent than are black men. Ratings of role congruency, or job-gendered context (and job-ethnic context), for this bookstore clerk job might be useful. These data could be collected from individuals other than your subjects in Study 2, so long as the sample was comparable.

(13) The Study 3 methods and results sections require some editing. (a) First, under employee race, last sentence, is the national average of 28% in reference to the % of the population, the working population, or the working population in this industry? (b) The objective facility characteristics section is repetitive of what appears in the specific sections (productivity, quality attributes) that follow. Also, the section on quality attributes is too detailed. (c) On page 32, the text incorrectly labels Tables 4 and 5 as Tables 5 and 6. (d) Check your variable labels in Tables 4 and 5 compared to each other as well as the methods section, as there are some inconsistencies. (e) In this study, you note that younger employees might be preferred, yet in Study 1 you noted that older employees might be preferred. A note about why this difference is expected is warranted. (f) At the bottom of page 31, you referred to an employee survey. What employee survey is this? It seemed that you had collected data from organizational records and customer-members, but never directly from employees.

(14) On page 33, several times you used the phrase “marginally significant.” According to null hypothesis significance testing, this does not exist. A result either is or is not significant; it is a dichotomous decision rule. Additionally, the p-value is not indicative of the strength of the relationship, so it is not to be used to describe the quality of a relationship between two variables.

(15) Figure 3 is not particularly illuminating. It doesn’t tell us anything new about the results, and as a summary it does not add to our understanding.

(16) The discussion seems rather short compared to the rest of the manuscript. Certainly some discussion occurred at the end of each Study summary, but additional thought about the constructs and the implications of this research for theory is needed. You did a fine job of describing the practical implications.

(17) On page 36, it might be inaccurate to refer to paying nonwhites and females less when they have lower customer ratings as “discrimination” (second line, p. 36). This is a very strong word that some people equate with intentionality to pay less.

(18) Throughout the introduction and discussion, you highlight your use of objective data as a strength. I agree, but I also believe that your statements should be somewhat tempered, as it is important to recognize that objective data are not automatically valid representations of the constructs of interest.

R2

The purpose of this paper was to examine whether customer satisfaction ratings are biased against women and minorities, and against facilities employing larger proportions of women and minorities. Using a very nice combination of three studies, the authors find powerful evidence that customers do indeed discriminate against women and minorities in their ratings. This paper has a number of real strengths: the authors were creative and thoughtful in their use of three different methodologies to test their hypotheses, the samples and data were quite impressive (especially for studies 1 and 3), the results were consistent and clear, and the paper was generally well written. There are, however, some questions/issues I have about the paper, and I briefly describe these below:

- 1) Your results, across all 3 studies, are quite powerful. Despite the strength of your findings, however, I felt that your discussion section was quite weak. Given the emphasis that AMJ places on identifying the implications of research, I urge you to work hard to bolster your discussion of both the theoretical and practical implications of your work. With regard to the practical implications, your suggestion was for organizations to stop using customer satisfaction ratings in personnel decisions. Although this may help reduce the perpetuation of bias against women and minorities, this is simply not a realistic suggestion! As you mentioned in the beginning of your paper, more and more organizations are relying on customer satisfaction data; thus, a more appropriate suggestion may have to do with how customers can be motivated to be less biased in their ratings (much like we do with employee raters), or how organizations can transform the ratings in some way to account for the existing bias against women and minorities. There may be other alternatives as well, and I urge you to think carefully about what they may be. As for theoretical implications, what do your findings tell us about the burgeoning literature on the diversity to performance relationship? And how does your work contribute to relevant literature on customer service organizations and service quality? In addition, you did not discuss the limitations of your studies (except briefly after study 1 in order to set up the reader for study 2). Please add a discussion of limitations.
- 2) Even though the overall story that you are able to tell with your 3 studies is powerful and consistent, I think that it is important for you to discuss how the characteristics of each of your study samples may have influenced your results. For example, when reading about study 1, I was struck by how women and minorities may be more likely to be evaluated negatively as physicians than they would as sales people due to greater perceived role-incongruence in the medical profession. Would you have seen such strong results if the employees being rated were sales personnel in a department store instead of physicians? You briefly mention this issue on page 19, but a more thorough discussion seems warranted. Similarly, in study 3, how might the country club environment influence customers' ratings of women and minorities, who have historically been excluded from such establishments? It is possible that there are greater biases against

women and minorities in country club settings than there might be in a YMCA, or in a run-of-the-mill restaurant chain?

3) **Study 1:**

- a. Given that both women and racial minorities are rated lower than equally performing men and Caucasians, respectively, I am curious to see whether minority women are at even greater risk due to their double-minority status. Are the ratings of minority women worse than those of minority men? There is increasing attention being paid to “double jeopardy” issues (e.g., Berdahl & Moore, 2006, JAP, 91, 426-436), and thus an examination of this issue may be worthwhile.
- b. It was difficult for me to follow your description of Relative Value Unites on p.14. An example that helps to illustrate this would be helpful.
- c. I was surprised to see that you did not control for patient gender and race. Although limited, there is some research that suggests that physicians provide better quality care to Caucasians than to ethnic minorities (e.g., Chen, Rathore, Radford, Wang & Krumholz, 2001 in the *New England Journal of Medicine*, 344: 1443-1449). Given this, minority patients may be less satisfied with their providers than white patients, and this could contaminate your results. Clearly, you seem to recognize the importance of rater race and age, as you controlled for it in Study 2. So why not in Study 1?

4) **Study 2:**

- a. Please expand on your discussion of how the perceived negative properties of a low-status employee can spill over into customers’ evaluations of the context on p.21. You mention two studies that loosely support this idea by showing that low-status employees damage the reputations of coworkers, but you do not provide evidence of how perceptions of employees can transfer to perceptions of the inanimate service context. Through what cognitive processes does this occur? Is there any prior evidence of such an effect? If not, then this is a real contribution to the service literature that you should emphasize more in your discussion. If this effect has been found before, then that literature should be duly acknowledged.
- b. Is the bookstore that was rated by students the bookstore of the university that they attended? If so, how can you be sure that their customer satisfaction responses weren’t influenced by their past experiences with the bookstore?
- c. I struggled to understand how respondents would be able to answer some of the employee performance questions based on the video scenarios that you described, and thus I wonder how appropriate the 6 items were as a manipulation check. For example, question 5 asks whether bookstore employees are too busy to respond to customer requests promptly. Yet I thought that both of the scenarios involve just one store employee, and since the videos were filmed before store hours, I assume there weren’t other store employees affecting respondents’ views? The other questions similarly ask about bookstore employees (plural), which I thought was confusing.

- d. The fact that you administered the race IAT before the second video is highly problematic, as there is no doubt that doing so primed your study respondents that your study objective had something to do with race. Please address this more completely by justifying why you administered the race IAT before the second video, and why you think this isn't a problem. If you do recognize that this is a significant problem, might it not be worthwhile to consider collecting additional data with your student sample (which is the easiest of your 3 samples to access) to rule out the possibility that this influenced your results?
 - e. I am intrigued by the fact that you did not find any evidence of rater biases when evaluating employee performance. My guess is that you attribute this to a social desirability effect (i.e., that respondents are more careful about their ratings in a lab setting) – that is, that people are careful about being unbiased in their ratings of video subjects because they don't want to be perceived as prejudiced. I wonder, if you simply examine people's ratings of the first video that they watched (i.e., before any IATs, and before people are aware that they are making two different evaluations), do you still find that there are no differences for ratings of minorities versus whites, women versus men?
 - f. As you do for studies 1 and 3, please plot the nature of the significant interactions so that the reader can see the slopes. Please also conduct simple slope analyses.
 - g. On p.26 at the beginning of the discussion section you say that you found that respondents rated employee and organizational context as being worse when observing the performance of a low status employee. But you actually didn't find differences for employees, right? This is a significant typo that needs to be fixed.
 - h. First sentence of the last paragraph before Study 3 on p.27 is unclear.
- 5) **Study 3:**
- a. What was the average response rate for each facility? An average of 63.8 customers out of an average of how many total? Also, only a percentage of each facility's customers were contacted with the customer satisfaction survey. On what basis were they selected?
 - b. The customer satisfaction scale seems to me like it might be multi-dimensional. Did you conduct a factor analysis to verify that it was unidimensional before creating a single scale score?
 - c. You indicated that 26.5% of employees were minorities, and 31% were women. Do you have data that speaks to how these individuals were distributed across jobs within the country clubs? That is, were they in role-congruent or role-incongruent jobs within the country clubs? What types of jobs were represented?
 - d. Controls: Considering facility profit margins are of interest, I would like to see you control for other factors that may affect profit margins. For example, member characteristics such as member age and socioeconomic status, and club factors such as centrality of location, classes/programs offered, restaurant and bar offerings, availability of other facilities such as

pools, etc. Any number of these factors could influence profit margins, and so you should control for them if you can get this data. Also, similar to my concerns regarding Study 1, I wonder why you didn't control for the race of customers in study 3. If you can get this data, please include it in your analyses.

6) Minor issues:

- a. On page 3, end of first paragraph: The sentence beginning with "From an organizational perspective..." is quite confusing. When I returned to this section after reading the complete paper I understood what you meant, but was very confused when I first read this. I think it is very important for you to be clear in this paragraph, as this is where you are selling the worth of your paper to the reader.
- b. Your theoretical rationale would be strengthened by the inclusion of more references about what motivates people to engage in less biased social cognition (i.e., to overcome stereotypic perceptions by collecting individuating information) on p.7. Fiske & Taylor's (1991) book entitled "Social cognition" may be a good place to start.
- c. The first sentence of your last paragraph on p.37 is confusing. What do you mean by "even when customers are relatively wealthy," and "characteristics designed to benefit them?"

R3

The current paper examines the effect of gender and race on customer's perception of the service they receive, as well as on the service context and the facility. The manuscript reports on three studies that examine whether and how biases may influence customer satisfaction ratings and produce discriminatory judgments for minorities and female employees. The current manuscript has many points of strength. First, although racial and gender bias have been widely researched, the current study addresses an important question attempting to better understand biases in the context of service interaction. Second, the study is very well designed in terms of its methods. It combines multiple methods of studies starting with a field study, continuing with an experimental study in the lab and concluding with another field study. The use of multiple methods and the ability to draw on the results and conclusions of each study, in order to develop the following study, enables the authors to reach a deeper understanding of the researched phenomenon. Third, this is a well organized and well-written paper. The hypotheses are clear, the methods and results sections are clear and easy to follow and the discussion section highlights eloquently the implications of the studies and the concerns they raise. Notwithstanding the above there are a few major concerns that require further attention. I explicate these below.

Abstract:

1. The abstract explains well the findings of the paper, however currently the authors do not note the problematic or unsettling nature of the findings of the study. Since abstracts are at many times seen without the full paper attached it is important to note the complexity of the findings in the abstract section. Thus, although the authors note this explicitly later on in the discussion of the study's results the complex nature of the findings should also be highlighted in the abstract. Therefore I recommend the authors state at the end of the abstract that the "troubling findings will be discussed in the paper" or note in any other way the complicated nature of the findings.

Theory and Hypotheses

2. Overall, I think one major limitation of the current manuscript is its potential contribution to theory development and to our thinking in the field of gender and race biases. A major emphasis of AMJ is contribution to theory building and theory development. This manuscript raises an important issue, however, currently it does not add new directions for understanding this bias. The racial and gender bias have been shown in many areas of research. However, this bias has not been well demonstrated in the employee-customer relationship. Furthermore, the study of the spillover of biases to the service context is a novel contribution of this paper. Thus, I think an important challenge of the current paper is to further think how the important research questions raised can be framed theoretically and how they can contribute not only to empirical knowledge, but also further our theory in this field. For example, the authors focus on the role of customers in the

- biases toward employees. Could the authors further develop a theory focused on the power customers can have to shape organizations, to shape rewards, selection processes, to influence management and thus lead to a systematic organizational race and gender oppression? Such a theoretical framework can be most interesting and novel.
3. Currently the paper is focused on studies that aim to test the existence of biases toward women and non-Whites. I believe a much stronger contribution could be made if the authors attempt to understand what can limit these biases. Currently the paper presents quite a gloomy picture, showing that things have not changed much and that biases are still prevalent. I believe a more meaningful contribution would be to suggest possible ways to eliminate these biases in the employee-customer relationship, and to test these predictions. Developing a theory that sketches new ways to reduce bias can be most helpful for scholars, as well as for organizations.
 4. Thus, the paper would be much more interesting and stronger had it provided ways that show how discrimination can be minimized. Maybe the authors can design another study that is focused on how to limit work place and customer biases. What would enable customers to see women and non-White employees as more competent in the service interaction?
 5. Although the paper is well written and well designed, I did not find the format of writing highly intriguing, as it was structured mostly around studies and not around theory. I believe this type of writing is somewhat more in line with the style of JAP or JPSP and is somewhat less attractive to AMJ readers. Furthermore, among the many references of the paper only one is from AMJ, further suggesting that currently this paper draws on a somewhat different type of stream of writing. I would suggest the authors further think on how the theory section can be re-written to put forward new directions of thinking possibly contrasting different theories or developing new theory in a format that will be more attractive to AMJ readers.
 6. p. 21 – The authors note that in the lab people will be less biased. This is in contrast to earlier studies which have showed the opposite. For example the meta-analysis of Eagly, et. al, (1990) showed that significant differences between men and women in leadership style were found in some cases in the lab context but not in the field. They explained their findings asserting that in the lab context there are less indication of what the expected behavior is and this leads to more stereotypical behaviors and attitudes of the participants in the studies in comparison to the field.

Methods and Results

7. As I stated above the studies are well planned and well described and presented. The authors should also be praised for collecting many relevant control variables.
8. Maybe it should be worthwhile to have control variables which relate to the customers. For example, in study 1, how often do the customers see this specific physician? Maybe if they see the physician more often they have a less biased attributions and attitudes toward the physician?
9. Could the authors obtain data regarding the ratings that managers and co-workers gave employees in the studied organizational contexts? If they obtain such data they can compare them to the ratings of customers? If they were to find that the managers and co-workers gave less biased ratings in comparison to customers they may be able to show that a longer exposure to the female and non-White employees may limit the gender and racial biases. Is it possible that customers that have long term interactions with the employees are less biased? All this should be further examined.
10. In the first study the n for non-whites is very small.
11. In the note for Table 2 – the authors should specify what IAT stands for.
12. Table 3 – please note what correlations are significant. Currently it is not marked.

Discussion

13. The discussion is well presented. I mostly found value in the section in which the authors explain that if women and non-White employees and their organizational units are seen as performing less well by customers, then these employees and units should receive fewer rewards, bonuses and promotional opportunities. I think this is a great point that may be the basis for a development of a theory of organizational level systematic oppression that starts from the bottom up. Starting from the customers and ultimately effecting the management. This could also be a basis for re-thinking the power of customers to shape the power relationships within organizations. I think that if the authors further develop these points in the theory section and in the discussion they can add to the depth and contribution of this paper.

A minor point:

14. The authors stress the novelty of their manuscript, while highlighting the disadvantages of prior studies. At times I thought that this line of reasoning did

not give enough honor to prior research. I would recommend the authors re-write these sections in a way that will point out and highlight the strengths and the contributions of the current paper, without playing down the contribution of prior work.