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

James G. Combs
Florida State University
jcombs@fsu.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
Boston College
amjpratt@bc.edu

Wm. Gerard Sanders
Rice University
gerry.sanders@rice.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

Feb 4, 2009

Dear Dr. Hekman:

On behalf of the entire editorial team, let me express our appreciation for submitting the revision of your paper to us. Titled "AN EXAMINATION OF WHETHER AND HOW RACIAL AND GENDER BIASES INFLUENCE CUSTOMER SATISFACTION RATINGS," this version of your work is logged into our system as AMJ-2008-0445-R1.

We sent your revision to the same three reviewers who provided feedback to you regarding the initial version of your paper. These scholars have completed their study of the revision. For your convenience, I am placing their reviews at the end of this decision letter.

As you will see, all three reviewers are quite pleased with the revision you have submitted to us. Indeed, all three reviewers feel that (as Reviewer 3 writes) "you have been highly responsive to all the comments raised by the Editor and the Reviewers." Moreover, all three reviewers indicate or suggest that the study makes an important contribution to the management literature and is likely to be highly cited. While there are still a number of issues the reviewers and I would like addressed before we send this manuscript off to press, most of these issues are, in the scheme of things, relatively minor and easily handled. As I will expand upon in a bit more detail below, the most significant reviewer concerns remaining at this point require that you: (a) better clarify and support some of the socio-cognitive mechanisms underlying the linkages suggested by the hypotheses (R1, point 1; R2, point 3; R3 – overview), (b) revise the phrasing of the hypotheses, (R1, point 3, R2, point 4), (c) adjust Figure 1 to better reflect the actual variables and relationships examined, (d) enhance the discussion section to provide a bit more detail regarding the potential solutions suggested (R1, points 1 & 2; R2, point 11; R3, point 1), and (e) expand the limitations section to cover a number of additional reviewer concerns that are either beyond the realm of the current analysis or simply cannot be addressed empirically given the nature of the data.

These critiques aside, it is obvious that you and your co-authors have worked hard to earn these positive reactions and to meaningfully address the issues the reviewers and I asked you to consider when preparing this revision. I too am pleased with the version of your work to which we are responding in this letter. I believe that you have effectively handled the core recommendations we presented to you. In light of our satisfaction with the work you have completed, and because of the consistent editorial recommendations I have received from the reviewers, I am pleased to **CONDITIONALLY ACCEPT** your paper for publication in Academy of Management Journal. Congratulations to you for earning this acceptance of your work! This is indeed a significant accomplishment. I hope that you are as pleased with this decision outcome as we are in reporting this decision to you. As you know, we are accepting your paper on the condition that you will successfully handle the issues the reviewers and I are presenting to you in the reviews appearing below and in this letter.

Before proceeding to express the concerns that I am asking you to address as you prepare what we hope will be the final revision of your work, let me provide information regarding the procedures we are asking you to follow to do this. I will offer these points next. As you will see, the materials appearing next deal with the handling of the revision process. This is important material for us to present to you even though you are familiar with these procedures.

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.

I should also note that we are trying to facilitate timely publication of manuscripts submitted to the Academy of Management Journal. As such, we wish to request that your revised manuscript be uploaded on the Manuscript Central system within the next two months. Please let us know if it is not possible for you to meet this deadline or if you feel that an extension beyond two months might be necessary.

With this information expressed to you, let's now turn our attention to the revisions that I would like to see you make. I will present a numbered set of points to which you should respond when preparing the revision.

1. Theory and contribution-related concerns: While there is little doubt about this study's empirical contribution, as suggested in my original letter, a strong theoretical contribution is contingent upon the clear and detailed explication of the dynamics underlying the hypothesized links. Although the current version now provides greater insight into these dynamics, all three reviewers feel that further explication and clarification are required. Based on their critiques and my own reading of the manuscript, I would like to see three main actions taken:
   a. Provide more insight into the socio-cognitive mechanisms that you believe are underlying the posited effects. As noted above, all three reviewers (R1, point 1; R2, point 3; R3 – overview) feel that by providing deeper insights into the mechanisms driving rater bias, the hypotheses could be made more compelling. A better understanding of the mechanisms is also important in order to assess the viability of the solutions you recommend in the discussion. Furthermore, I would like to see a stronger link back to theory and empirical research as you discuss these mechanisms. There are two main aspects of the theory that I see as especially requiring such action:
      i. Anonymity/Accountability dynamic: While I like the reference to the anonymity and accountability literature, as noted by Reviewer 1 (point 1), there is an important difference between anonymity

and accountability. Might there be something about non-anonymous rating that affects rater behavior even when raters cannot be made formally accountable? In my own research on peer rating, I have found the game theory literature, and in particular, the notion of "backwards induction" to provide important insights into such questions (see Bamberger, "Competitive Appraising: A social dilemma perspective on the conditions in which multi-round peer evaluation may result in counter-productive team dynamics," *Human Resource Management Review* 17(1),1-18).   You may also wish to look at the research on rater accountability by David Antonioni (e.g., *Personnel Psych*., 2006) in this regard.

    ii. Spillover dynamic: Additionally, as noted by Reviewer #2 (point 3), the theory underlying the spillover effects suggested in H2 remain less than compelling. In particular, you never really make clear whether this spillover is conscious/intentional or unintentional on the part of the rater. This reviewer suggests that you consider connectionist theories to help in this regard.

  b. Rephrase the hypotheses: As currently phrased, several of the hypotheses lack clarity and may even be considered less than falsifiable.   Reviewers 1 (point 3) and 2 (point 4) offer a couple of good suggestions in this regard. If I might add my own two cents, you might want to consider replacing the terms "less (more) positive" with notions of attenuation and amplification.

  c. Figure 1: All 3 reviewers find Figure 1 to be rather confusing as it references constructs and terms that are only peripherally mentioned and not among the constructs incorporated in the hypotheses. Furthermore, it suggests patterns of relationships other than those hypothesized or tested in your analysis. For example, while the "task attributes" are assumed to underlie many of the observed effects, because they are not directly tested. Accordingly, perhaps they should be excluded from this figure, or if included, graphically shown to be outside of the boundaries of the model which you actually test.

2. Discussion: The reviewers and I are intrigued by your discussion of the potential solutions to the customer bias problem documented in this study. That in mind, they (R1, points 1 and 2; R2, point 11; R3, point 1) ask for further clarification of some of approaches that you recommend be given consideration by managers. In particular, while recognizing space limitations, the reviewers and I would like to see a bit more discussion of any empirical research on the application of such approaches with respect to customer ratings, and an analysis of the advantages and disadvantages of these alternative approaches. For example, has any research been conducted on the application of the approach used in the Olympics (to drop outlier judgments) to consumer satisfaction ratings? Additionally, in terms of accountability, while customers can rarely be made accountable to those they rate, at least in those contexts in which there is likely to be repeated interaction (e.g., with one's physician or golf pro), backwards induction suggests that non-

anonymity may indeed be effective in enhancing rater accuracy (i.e., reducing bias).

3. Limitations. The reviewers also suggest a number of study limitations not mentioned in the text, and particularly those relating to design flaws in study 2 (R1, point 12), concerns with regard to the IAT (R1, point 5; R2, point 10), and remaining questions regarding the suppression effect and rater demography issues raised in the first review (R1, point 20; R3, point 2). I recognize that you addressed several of these concerns in the letter accompanying your revision (e.g., your treatment of my original points 2, 3a, 3bii). However, since these issues continue to raise concern among the reviewers, I suggest that you make mention of these post-hoc tests in the text itself. You need not go into detail, but I think it is important that these threats to internal validity be recognized and, to the best of your ability given data limitations, addressed. I suggest placing this discussion in the limitations section in that, as you acknowledge in your letter, your data do not always allow you to fully rule out these threats. In that respect, despite your post-hoc results, several of these concerns remain and indeed serve as valid study limitations that need to be addressed in future research.

4. Explanation of Control Variables: I agree with Reviewer 1 (points 16 and 17) that the justification for the inclusion of several of the control variables is less than compelling, and – in some cases – less than "P.C." That in mind, I concur with the Reviewer's recommendation that the explanation be dropped. In addition to the two controls mentioned by the Reviewer, I had questions about the logic underling the "temporary employees" control in the Golfcorp study: Couldn't the argument be made that temporary employees exert greater effort than permanent ones in the hope that they might be offered permanent employment? Indeed, the correlation of this variable with satisfaction (+.15) suggests that the relationship is NOT inverse as suggested in the methods section.

5. Manuscript Length. The text is now 47 pages in length (not including tables and figures). I agree with Reviewer 2 (point 1) than 3-5 pages could be easily eliminated, bringing the manuscript to within range of the high-end of "normal" for papers published in *AMJ*. This Reviewer offers some excellent suggestions as to how this might be accomplished and I encourage you to follow them. In addition, I believe that additional cuts may be made by: (a) condensing your justification for the inclusion of specific control variables (as noted above), and (b) reducing the number of plots of sample interactive effects (since many depict a similar pattern). The latter two cuts should allow for an *additional* reduction of 2 pages in TOTAL manuscript length.

Although I am going to stop here, please take a look at the other suggestions of the reviewers. While I see most of these as less critical, by taking these recommendations into consideration you are likely to significantly enhance the "final polish" of your manuscript.

Overall, as I hope you will see, while we are requesting a significant number of revisions from you, most of them are relatively "easy fixes." Again though, each of the actions the reviewers and I are asking you to take is important and has the potential to meaningfully enhance your work. Thus, as noted above, please devote careful attention to dealing with the requests we are making of you.

Congratulations again to you for earning the conditional acceptance of your work by Academy of Management Journal. And, thank you for submitting your work to AMJ for publication consideration. We appreciate the opportunity to work with you and truly look forward to receiving the next revision of your paper (which should be coded as AMJ-2008-0445.R.2).

Warm regards,

Peter

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

R1

This revised manuscript examines a full-cycle of studies on ethnicity- and sex-biased judgments of customer service. As with the initial submission, this is a strong manuscript that is interesting to read and has a number of strengths. I also felt that the authors were responsive to reviewer concerns, and the paper is considerably stronger now. I have a number of comments regarding the clarity of the work and additional changes that could strengthen the paper.

(1) On page 8, you wrote about the effects of anonymity and accountability on customer's motivations to rate without bias and how supervisory ratings differ. You should clarify that supervisor's ratings are (or can be) reviewed by their supervisors and/or HR, and they can be held accountable for their ratings. You've implied this, but being more explicit would be better. Then, note that although customer ratings are also reviewed by the same agents, the customers are not accountable to these people.

(2) On page 9, you wrote about the different components of the ratings forms that contribute to bias. You have glossed over the lack of BARS in this section. Bolster this issue. It is important because it may be that customers aren't "biased in ratings," per se, but rather they are "biased in expectations." (e.g., expect greater warmth from a female doctor than a male one, so when they get equal warmth from men and women, the women do not seem warm enough). Including BARS might help alleviate this problem. I also think this would make a good discussion point.

(3) All of your hypotheses have a flaw in the way that they are stated. In particular, you say that there is a "less positive" relationship. This is ambiguous, as it could mean that there is a main effect or an interaction effect (which is what I think you mean, depending on the hypothesis) OR that there is a difference in the correlation (for example, comparing a correlation of +.80 with one of +.25). Although in a sense there is no difference, one issue is that with the latter interpretation regarding differences in correlations, it implies that the relationship is less STRONG, not less POSITIVE (i.e., shift in position of the regression line), because the correlation is indicative of the variance in one variable accounted for by another. Maybe you do mean this, but it just isn't clear as the hypotheses are stated. Additionally, the latter interpretation doesn't seem to be consistent with your theoretical positions (see, for example, page 10, lines 34-38).

(4) On page 11, I was curious, regarding the Argo et al. (2009) study, was this effect of an attractive person of the opposite sex occurring only for heterosexuals? If so, you should note it.

(5) I still have some issues with the way you frame your use of the IAT. First, you really don't acknowledge the debate over the meaning of IATs (e.g., Blanton & Jaccard, 2006). Second, on page 13 (lines 13-18), you noted that you tested effects of non-conscious, negative attitudes toward low status groups. Yet the preceeding arguments on pages 12-13 seem to suggest that although most of us have negative racialized and sexualized thoughts (which are nearly unavoidable due to the socio-historic biases in the United

States, and elsewhere—but for different groups), our internal commitments to not being biased can suppress these automatic thoughts. Be sure that you clarify which you are talking about, what it means, etc. I don't dispute that you have something interesting happening. But the IAT is not a "complete" measure of "racism" or "sexism" and without extremely careful attention to wording, individuals who are less familiar with this line of research can easily misconstrue the results and implications of your work.

(6) Hypothesis 3 should clarify whether you're referring to a greater proportion of women and minorities, or a greater number.

(7) Figure 1 depicts vulnerability to bias as a mediator, yet your hypothesis and your theory seem to depict it as a moderator. Clarify and be consistent.

(8) Regarding physician accessibility (page 19), if patients can send unlimited numbers of emails, isn't the more appropriate measure the percent of emails replied to within the organizational expectations of 24 hours, rather than the total number of emails sent? Also, how do the doctors decide who to be accessible to (line 37, page 19)? Is this about replying to emails, or blocking access in the first place? What if no one emails you (e.g., your panel is older), or you have a smaller panel than other doctors (e.g., you work fewer hours)? Yes, these are some things that are controlled for, but they are controlled in the regression of customer ratings, and not accounted for in the differences in the predictor variables.

(9) Relatedly, doesn't age of the panel influence the prescription rate (physician quality—which is a misnomer, really—prescription rate is probably better)? Again, panel age is controlled in predicting customer satisfaction, but is not accounted for in the base rate of the predictors. (Ostroff has a great paper that demonstrates how differences in means of a variable across cases/groups can influence the regression line in a way that obscures the true relationship.)

(10) Despite the fact (as you noted in your response letter) that similar findings appear in another AMJ paper, isn't it odd (and possibly contrary to your hypotheses) that your interactions show that better performance is more negatively rated for women and minorities? A flat slope, or a barely increasing one should be expected given your theoretical positions, not a negative slope. This is surprising, and should be explained further.

(11) In the Medcorp and Golfcorp samples, is it possible to report the N of employees who are both female and minority? I recognize that the samples are too small, or the variable to participant ratio too skewed, to do analyses. But if all ethnic minorities are female (for example), this might be important to note.

(12) The Bookcorp experiment is flawed for not including an African-American female target. Also, your logic for not including this condition in the reviewer response letter belied the logic of full cycle research, essentially saying that you did the two field studies and then the lab, rather than field-lab-field. If you only had a series of studies like the

Bookcorp study, it would (in my opinion) be "fatally flawed." However, given the totality of your project, I suggest that you provide a stronger acknowledgment of this flaw and emphasize the importance of including this condition in future research.

(13) page 28, line 13—not sure it is correct to say that you had two "conditions." You had three (white male, Af-Am male, white female), and have set up your analysis as two "mini-studies" embedded in the same paradigm, using the same control group. You might consider redoing your analyses so that you have two dummy variables (0/1 white male & white female/Af-Am male; 0/1 white male & Af-Am male/white female) so you can actually compare across the two low status conditions and graph them all together. I think it would be more interesting this way.

(14) The surprising results (flat line for white men, X shape of interaction) for the first panel in Figure 3 needs greater attention in the text.

(15) page 32, line 13: your statement that the organizational context is distal from low status employees is inconsistent with your introduction (and the sentence that follows this statement) in which you note that there is a fuzzy boundary between figure and ground, or people and people around them.

(16) Wouldn't larger facilities have more amenities, thus making them more likely to be satisfying?

(17) I don't buy your footnote 4. Why would younger workers be more eager to please? What evidence supports this contention? I recommend just stating that you controlled for this. Also, I found your statement that male customers would prefer the club because it is a golf club to be extremely sexist in the absence of supporting evidence. Just note that you are controlling for percentage of male customers. Finally, longer tenured members not only are "more familiar" with the club, but also (probably) more satisfied, because they haven't quite the club (which, of course, they can do).

(18) In Tales 5 and 6, the signs reverse for sex and ethnic minorities as predictors (from the correlation table to the regression table). Thoughts?

(19) On pages 44-45, you note some of the additional control variables that were added and struck. There is too much detail here, especially because if you violate the rule of thumb regarding cases-to-variables ratios, then you can't take the results too seriously anyway.

(20) I noted this in my initial review, and I will express my concerns again: there seems to be a suppressor effect in your analyses for Medcorp. In particular, there is a sign reversal for panel age and age of physician, which have a non-trivial relationship. You should investigate this. I was not persuaded by your response letter. This issue should be acknowledged in the manuscript.

Minor changes:

(a) page 6, lines 32-37: delete the sentence about Obama and Clinton. It's timely only today, and the fact is that there are a lot of reasons why they were supported now and why others weren't in the past (not just changing attitudes). Plus, it's distracting from the flow of the paper.

(b) page 7, line 44: switch to "men and whites" to be parallel with "women and ethnic minorities" in the next line.

(c) page 8, line 25: the correct spelling is Drasgow, not Drascow.

(d) page 9, line 41: correct spelling is DeNisi, not Denisi.

(e) page 9, line 51: change "will not be observed" to "are not present." The former could imply that the effect is present but not able to be detected.

(f) page 15, line 14: the header "Methods" doesn't really fit here. Delete and note that "this paper reports tests of the hypotheses in our model…" Similarly, don't refer to the three studies as "samples." They are "studies." So change the headers to Medcorp Study (or Study 1: Medcorp), etc. This is more consistent with the kind of research you're doing, which is not straight replication in different samples. You might also note here that all three organizational names are pseudonyms, and delete this footnote from the Golfcorp study.

(g) page 25, line 27: I would bet cash money that your student sample is pretty darn familiar with doctors too, not just bookstore employees.

(h) page 34, line 6: state that it is the "US population".

(i) pages 35 and 36 (and elsewhere in Golfcorp sample): use either "non-temporary" or "core," rather than "permanent" to refer to workers who are not temps. This is consistent with the contingent work literature (e.g., Hulin & Glomb, 1999).

(j) page 35, line 51 (and relevant tables): your variable is actually coded as percentage of NON-temporary workers, and should be named as such.

(k) Figures 2, 3, and 4 should have the panels re-ordered so that they follow the order of the description in the text. Also, go back and double check Tables and text (as well as across tables) so that all variables have consistent names. (e.g., page 22, line 44: is "responsiveness" the same as "accessibility to patients" in Table 1? Table 5's "Quality of Capital" is the same as Table 6's "Facility Quality"?)

R2

Once again, I really enjoyed your paper. The combination of the three studies is impressive, and I felt that this revised version of the paper was stronger than the first. In particular, you did a great job of strengthening the discussion section. Your discussion of the practical implications of the study for women and racial minorities was compelling. I also enjoyed the tone of your writing, especially how you provided rich detail and context about your studies and also about the analytic decisions that you made along the way. Below are several additoinal issues for you to consider:

1) I realize that I just said that I loved the detail that you provided in your paper, but nevertheless, the paper is a bit too long. I of course realize this is because you are reporting 3 studies, but it would benefit your readers if you could shave 3-5 pages from your text. You can start by deleting the paragraph on full-cycle research on p.39 – your description of this in the introduction is great and there's no need to repeat it. Also, the second and third sentences under "theoretical background and hypotheses" can be deleted, as they repeat what you've already described in the excellent introduction. Finally, the paragraph that comes after "Insert Figure 1 here" on p.14 can be deleted. At this point, I remember what your hypotheses are and don't need you to remind me. Other than that, I think that you can cut length by going through the manuscript sentence by sentence and trying to identify ways to say the same thing in a more efficient manner.

2) Your reference to the Stasser & Steward (1992) study about murder mystery games on p.8-9 seems out of place for the argument you are trying to make about rating accuracy. I see the link you are trying to make, but customers aren't trying to "find a solution" per se, so I would prefer to see you exclude this reference. Similarly, your discussion of religious symbols and good luck charms on p.11 feels like a stretch. You have enough relevant organizational support without including this.

3) The theoretical discussion that you provided for why rating inaccuracy might occur is much stronger now than it was in the last version of the paper. One section that could still be strengthened, however, is the one leading up to your spill-over hypothesis (H2).

   a. While you reference a good set of articles in developing hypothesis 2, you still do not provide a description of the mechanism(s) through which spill-over occurs. What theories do the spill-over studies that you site rely on to explain this effect? What social cognitive processes may be at work here? You may find it useful to explore connectionist theories, since customers' schemas of service employees and the service context would be connected schemas within a larger cognitive network of information having to do with the particular type of service experience; thus, evaluation of one would be intertwined or "connected" with the evaluation of another.

   b. Your discussion of this spill-over in the first full paragraph on p.12 makes it sound like the spill-over is the result of a conscious decision-making process – is this your intention? Perhaps for some customers who are quite prejudiced, the process is conscious, but for many, I suspect this

spill-over occurs at a more unconscious level of processing. Whatever your stance on this, it needs to be clarified. Perhaps both are possible, but you need to state this and support your arguments accordingly.

4) Your Hypothesis 3 needs to be split into two separate ones (H3a and H3b) because you are not testing the moderating effect of negative attitudes on the relationship between service employee race/gender and customer satisfaction with the employee, and that between service employee race/gender and customer satisfaction with the context simultaneously.

5) I am not a big fan of Figure 1. I found it to be confusing. First, the box labeled "task attributes" is confusing – what task? I realize you are talking about the rating "task" but your choice of the term makes it sound like a job task. Regardless, I don't think that this should be in the figure because it is not part of the model that you are actually testing. Second, the label "vulnerability to racial/gender bias" is misleading – why not just say "race/gender of employee" since this is what you are actually measuring. When I first looked at this figure, I thought you were introducing a new construct that you had not yet described in your paper. Third, rather than say "judgment of…" if you labeled these as "customer ratings of.." it would be more consistent with the terminology that you use in the text and therefore easier to follow. Finally, I understand conceptually why you have an arrow from "judgment of context" to "judgment of organizational unit," but I spent some time trying to figure out where you test this and realized that you don't. Perhaps you can exclude it or denote it with a dashed line?

6) I think it would be helpful for you to foreshadow before the methods section which hypotheses are going to be tested in which studies. Earlier I suggested that you delete the paragraph summarizing your hypotheses on p.14. You can use this space instead to *briefly* describe which hypotheses are tested in which studies. Indicating this in your figure would also be helpful.

7) For the Medcorp sample, do you have any data indicating how long patients have been made to wait? My thought was that in addition to average practice busyness, this could be an important control variable. Suppose that, consistent with research which shows that women are more empathic, female doctors spend more time with their patients. As a result, their successive patients will have to wait longer. Even though these patients might get good objective care, the may be frustrated because they had to wait (and this gets reflected in the customer satisfaction ratings).

8) In your description of the simple slopes (particularly for the medcorp sample, p.23 and p.24), it is not necessary to provide statistics showing that the slopes are significantly different from each other. This is obvious, since this is the reason why this step involving the interaction is significant in your regression model. However, keeping your sentence about how the simple slopes are not significantly different from zero but are still significantly different from each other is important.

9) I know that biases depressed customer satisfaction ratings for minority and female physicians based on the results you present, but can you also give us a sense for

whether there were any significant differences in quality, productivity, or accessibility?

10) I'm still not convinced that administering the race IAT *before* having subjects rate the videos was not problematic. At the very least, you should justify why you did it this way for race but not for gender, and explain why you don't think this is what accounted for your non-significant effect involving race, for hypothesis 1. You state on p.28 that "there is no evidence of order effects for the IAT and dependent variables" – are you talking about other research or yours? You can't say that about yours, as you have no way of comparing people in the race condition who took the IAT before and after their ratings. Although responses to the IAT can't be faked, taking a race IAT makes it quite obvious that you are interested in something having to do with race. Thus, your subjects are primed, and this may be why H1 is not supported. You touch on this a little bit in your discussion section when you talk about how covert biases are more likely to be manifested in ratings of contexts, but I think you should also include something in your discussion for Study 2.

11) I really like the suggestion that behaviorally anchored ratings scales may be one way of reducing customer biases. Is there any research on their use with customers? Again, great job with the discussion section!

R3

The authors have made substantial modifications to the initial submission. The authors have been highly responsive to all the comments raised by the Editor and the Reviewers. The revised version of the paper has improved in terms of anchoring and clarifying the theoretical assertions and hypotheses, as well as in advancing their analysis. I think the manuscript on the whole makes an important and novel contribution. Nevertheless, I still have some concerns with the revised version and I will elaborate on these concerns.

My major concern is related to the theoretical underpinnings of the paper. I still think that the theoretical perspective of the paper is not as developed as it could be. I think the authors should further develop other mechanisms that my effect the customers' ratings apart from the three task attributes suggested.

Below I give specific comments, according to the pages of the revised manuscript, that indicate where I believe the authors could strengthen the paper:

Theoretical background:

1. The techniques to reduce customer ratings biases that are presented in p. 10 need to be better explained. The authors state that: "A variety of techniques have been shown to effectively reduce bias in performance appraisal. For example, Roch and O'Sullivan (2003) found that a combination of frame-of-reference (FOR) and behavioral observation training (BOT) leads to increased accuracy in appraisal. Baltes and Parker (2000) found that halo error training and structured recall memory intervention reduces biasing performance outcome feedback on ratings." There is a need to better explain what these techniques are.

2. Reading the theory section there are various issues that are not clear. First, the authors do not relate to the gender and race of the raters. Does it effect the ratings? Is there an issue of similarity? Do people rate better people who resemble them? Are the customers similar to the individuals being rated?
Second, Hypotheses 3 asserts that: *the judges negative attitudes toward members of low-status will moderate the relationship between employee status, context and customer satisfaction. Based on this Hypothesis and* following the same lines I would expect that after Hypothesis 4 the authors will offer Hypothesis 5 that relates *the judges negative attitudes toward members of low-status with the percentages of employees belonging to low-status demographic groups (i.e., women and minorities) compared to units that employ higher percentages of employees belonging to high-status demographic groups (i.e., men and whites).* Currently it seems as though the theoretical model and hypothesis are based on the data that was collected and there are theoretical holes that are not addressed. If the authors can not address these issues in the current manuscript this should be noted clearly as a part of the model that can not be tested.

Methods and Results

In the theory section the authors contend that the customers may give biased ratings, since they rate on general performance and not on specific behaviors. However, in the MEDCORP sample the ratings for customer satisfaction are based on specific behaviors: "Customer satisfaction. …..Patients rated each of the three items targeting their physician on a 5-point Likert scale (1=very poor; 5=excellent) "How would you rate…" (1) the attention the provider paid to you; (2) this provider's thoroughness and competence, and; (3) your opportunity to ask questions of this provider. The three items were highly correlated (average correlation is .93) so the organization combined them to create a composite patient satisfaction variable. The organization did not provide us with access to raw patient-level surveys.

Looking at Figures 2-4 I find it hard to understand the interactions presented. For example, in Fig. 2 (the first one on the left): Among non-white physician when the 'Physician Objective Quality' is low customers are more satisfied with the Physician than when the 'Physician Objective Quality' is high? This is hard to understand. Why would they be less satisfied? This pattern is evident in many of the interactions. Could the authors explain this?

Figure 1 includes the Task attributes which are not part of the model being tested in the studies. It should be acknowledged in the Figure that this part of the figure will not be tested.

I wish the authors good luck with thus interesting work!!!