AN EXAMINATION OF WHETHER AND HOW RACIAL AND GENDER BIASES

INFLUENCE CUSTOMER SATISFACTION

David R. Hekman
Sheldon B. Lubar School of Business
University of Wisconsin-Milwaukee
414.229.6296
hekman@uwm.edu

Karl Aquino
Sauder School of Business
University of British Columbia

Brad Owens
Terence R. Mitchell
Michael G. Foster School of Business
University of Washington

Pauline Schilpzand
Warrington College of Business Administration
University of Florida

Keith Leavitt
Michael G. Foster School of Business
University of Washington

AN EXAMINATION OF WHETHER AND HOW RACIAL AND GENDER BIASES

INFLUENCE CUSTOMER SATISFACTION[1]

ABSTRACT

We examine whether and how various biases may influence customers' satisfaction evaluations and produce discriminatory judgments for minorities and female service employees. We argue that customer satisfaction evaluations are biased because they are anonymous judgments by untrained raters that usually lack an evaluation standard. In our laboratory and field samples, we find disturbing evidence generally confirming our arguments and suggesting that the presence of nonwhite and women service employees may produce lower aggregated customer satisfaction evaluations which may ultimately hurt individuals and organizations financially. (84 words)

Customer satisfaction surveys have become a common source of performance feedback for employees and organizations (Hagan, Konopaske, Bernardin, & Tyler, 2006). Mercer Consulting Group reports that in 2006 customer satisfaction surveys were of primary importance for strategic decision making and over two-thirds of organizations used such surveys to determine some aspect of employee compensation (Mercer Consulting, US Policies and Practices Report, 2007). Moreover, customer satisfaction is an important predictor of a wide range of financial measures (see Gupta & Zeithaml, 2006 for a review) so it is not surprising that some companies tie some portion of employee compensation directly to customer satisfaction. For example, a one percent change in customer satisfaction for an average Fortune 500 firm has been shown to lead to a 1.02% change in *Tobin's q* which equates to a change of $275 million in firm value (Anderson, Fornell & Mazvancheryl, 2004), a $55 million gain or loss in cash flow in the next year (Gruca & Rego, 2005), and a 5.03% change in Return on Investment (Anderson & Mittal, 2000). Satisfying the customer is also becoming increasingly important to organizations as the global economy becomes more service-oriented. Macroeconomic trends indicate 76 percent of U.S. employees work in a service industry, and by 2016 the number of employees working in a service industry is expected to increase by over 17 million (Figueroa & Woods, 2007). The expanding service sector is perhaps one reason why among 681 senior executives surveyed by *The Economist Intelligence Unit* during October–December 2002, 65% reported customers as their main focus over the next three years compared to only 18% who reported shareholders as their main focus.

Many business leaders (Bracken, Church & Timmreck, 2001) and researchers (Salam, Cox & Sims, 1997) have applauded the use of customer satisfaction surveys because they believe that aggregated evaluations are highly reliable measures of employee performance quality. However, a potential disadvantage of using customer surveys, particularly for purposes of making compensation or

promotion decisions, is that they are ultimately subjective *judgments* and, as a result, are vulnerable to well known biases, including those based on the bandwagon effect, confirmation of pre-existing beliefs, education or cognitive ability, as well as stereotypic biases based on the race or gender of the person being rated (Gilovich, Griffin & Kahneman, 2002). Some researchers suspect that biases are unavoidable when gathering subjective evaluations of performance, especially when such judgments come from naïve and inexperienced raters who are not held accountable for the accuracy of their ratings (Pulakos, White, Oppler & Borman, 1989; Wilkinson & Fontaine, 2002; Woehr & Roch, 1996) To date, though, surprisingly little research has examined this possibility by empirically testing whether different biases influence customer judgments about whether an organization's employees, organizational attributes, or services meet or exceed their expectations.

The purpose of our research is to examine whether and how customer satisfaction ratings are potentially influenced by gender and racial bias. We extend the existing literature on biases in supervisory ratings of employee performance by focusing on *customer satisfaction* ratings, which have been mostly excluded from organizational behavior research (see Moshavi, 2004 for a rare exception). We conceptualize these satisfaction ratings as judgments and we examine not only judgments made about individual employees, but also about the organizational context (e.g., perceived cleanliness or appearance) and the organizational unit as a whole. To our knowledge, no research has examined bias in customer judgments of the organizational context or the overall organizational unit. Based on the literature on modern forms of racial and gender bias (Greenwald & Banaji, 1995; Crandall & Eshelman, 2003), we contend that customer satisfaction with organizational contexts or units may also be vulnerable to such biases. Finding empirical evidence of racial or gender biases in customer satisfaction would suggest that, from the organization's perspective, there may be a financial incentive

to favor white or male employees. Such a finding may help explain the persistent inequality between demographic groups in the workplace.

Two methodological attributes of our research distinguish it from previous studies of bias in customer judgments of employee performance and allow us to conduct a stronger test of the validity of our predictions. First, we take into account employees' *objectively-measured* performance behaviors when examining customer judgments. The problem with relying solely on customer satisfaction scores to assess customer bias is that such scores can be interpreted as capturing both true performance scores and biases (Latham & Wexley, 1977; Landy, Shankster, & Kohler, 1994). As Rotundo and Sackett (1999: 816) summarize this state of affairs, "There is no definitive way of determining whether the rated criterion used in a validity study is biased. Thus, there is no current method of establishing whether there is bias in performance ratings." Our study design allows us to tease apart differences in satisfaction judgments that are attributable to customer bias arising from employee demographic characteristics from those that are due to objective employee performance (Greenhaus, Parasuraman & Wormley, 1990; Pulakos et al., 1989; Wilkinson & Fontaine, 2002; Woehr & Roch, 1996).

Second, we heed recent calls for researchers to engage in "full-cycle" research where initial field-based findings are tested in the laboratory and then re-validated in a different field setting (Chatman & Flynn, 2005; Cialdini, 1995). Full-cycle research allows researchers to compensate for the weaknesses of one context or study design with the strengths of another. It also allows the researcher to investigate a broad initial question in a field setting (e.g., is there bias in customer judgments?) followed by a laboratory study that can utilize more control and enable the researcher to examine more specific questions in detail (e.g., what is a potential cause and consequence of bias in customer judgments?). Finally, the investigator can move back to a field setting to confirm findings

found in the first two studies. The interplay of field and lab designs prescribed by the full-cycle approach fosters greater theoretical insight as to the causality and generalizability of study findings.

Following the full-cycle research model, we first test for bias in customer judgments regarding a sample of professional employees (e.g., doctors). Next, in a carefully-controlled laboratory setting, we test for customer bias again, but this time in a bookstore context. We also identify and measure a specific mechanism that might explain the observed effect. Finally, we test for customer bias in judgments of an organizational unit with a sample consisting of country clubs belonging to a large hospitality company. Since the focal unit rated shifts from the individual in the first sample to the organizational in the third sample, we are able to provide some initial confirmation of the generalizability of our theory. In both field samples, customer ratings were used to determine employee salaries, which meant that any biases in customer satisfaction could later translate directly into differences in pay and the generation of gender and racial inequality in other career outcomes. In the following section, we present the theoretical rationale for our predictions regarding the possible effects of customer diversity related biases on customer satisfaction judgments.

THEORETICAL BACKGROUND AND HYPOTHESES

An assumption of our paper is that like anyone else who makes a social judgment customers are not immune to information processing biases. When customer satisfaction judgments are used to determine employee outcomes or as a basis for evaluating organizational performance, these biases can disadvantage some employees or organizations if the judgments are systematically lower than would be warranted based on more objective indicators of service quality. In practice, customer satisfaction judgments can be used to assess different aspects of the customer exchange experience. Most common are judgments of an individual service provider (e.g., their salesperson, teller, teacher, or physician; Haas, Cook, Puopolo, Burstin, Cleary & Brennan 2000; Sixma, Spreeuwenberg, & van der Pasch,

1998; Davis & Davis, 1999). Customers can also be asked to rate the quality of the environment in which they were served (e.g., the merchandise available, the newness or cleanliness of the setting, or the efficiency of the technology; Pellegrin, Stuart, Maree, Frueh, & Ballenger, 2001; Simonet, 2005). Finally, customers are sometimes asked to give an overall judgment of the unit or group providing the service (e.g., the bank, school, country club or law firm; Anderson et al., 2004; Ittner & Larcker, 1998). This last type of judgment is likely to include opinions about both the server(s) and the context in which the economic transaction occurs and is therefore a more global judgment than the other two types previously mentioned. In this study, we investigate the possibility of systematic bias in all three types of satisfaction judgments thereby providing a strong test of the potential generality of such biases.

Racial and Gender Biases in Customer Satisfaction Judgments

U.S. society has made considerable progress in reducing overt expressions of prejudice since the Civil Rights movement during the 1960s (Bobo, 1998). The recent election of Barack Obama as President of the United States and the willingness of millions of Americans to support Senator Hillary Rodham Clinton as a viable candidate for the same office testifies to this change in American racial and gender attitudes. Yet despite the gains made in reducing overt expressions of prejudice, there is abundant social psychological evidence that biases against women and minorities persist in a more covert and non-conscious form. Researchers have used terms like modern racism (McConahay, 1983), aversive racism and sexism (Dovidio & Gaertner, 1981; Gaertner & Dovidio, 1986), or implicit gender and racial stereotypes (Greenwald & Banaji, 1995) to describe these types of biases and many studies have demonstrated their influence on information processing and judgment across a variety of social domains (see Brief, Dietz, Cohen, Pugh & Vaslow, 2000 for an overview). For example, one study showed that when job applicant resume quality was ambiguous, applicants with African-American-

sounding names (e.g., Aisha, Rasheed) were much less likely to be called for a job interview than applicants with white-sounding names (e.g., Kristin, Brad) (Bertrand & Mullaninathan, 2004). Similarly, when evaluators of orchestral position applicants could see the applicant's gender they were more likely to select men. When the applicant's gender could not be observed, the number of women hired significantly increased (Goldin & Rouse, 2000). As a third example of the operation of covert biases, Dovidio and Gaertner (2000) found that while raters were not biased against blacks in a simulated hiring decision when the applicants were clearly qualified or unqualified for a job, raters were biased when the applicant's qualifications were ambiguous. Dovidio and Gaertner (2000) interpreted this finding as supporting an aversive racism framework in which prejudice occurs in a more subtle form in contexts with ambiguity or uncertainty. Based on considerable evidence demonstrating the operation of covert and unconscious racial and gender biases across a variety of social domains, there is reason to suspect that such biases can also influence customer satisfaction judgments.

Our theoretical arguments supporting the influence of racial and gender biases in customer satisfaction judgments are based on the idea that observers (e.g., customers) have preconceived notions about others depending on whether the person being observed belongs to a high or low-status demographic group (Berger, Conner & Fisek, 1974; Berger, Fisek, Norman & Zelditch, 1977). In the United States, whites and men are considered by most people as members of a high-status social group relative to women and ethnic minorities (see Ridgeway, 1991 for a review). One of the benefits of belonging to a high-status social group is that observers are more likely to make favorable inferences about one's competence, normality, and legitimacy (Aquino & Bommer, 2003; Giannopoulos, Conway & Mendelson, 2005; Sidanius & Pratto, 1999). In contrast, members of low-status groups are subject

to negative stereotypes and attributions concerning their work-related competencies (Fernandez, 1981; O'Leary & Ickovics, 1992).

Based on rating theory (Wherry & Bartlett, 1982), we believe there are at least three reasons to expect customer satisfaction ratings to be susceptible to such stereotypes and racial and gender biases. First, an important difference between performance ratings made by supervisors and judgments made by customers is that the latter are afforded the luxury of anonymity. Customers almost never have to identify themselves when they complete satisfaction surveys and so they remain unknown to the organization and to the service providers they are judging. This anonymity makes them less accountable for their judgments and decreases the likelihood that they will respond to questions in a socially appropriate way (i.e., suppress their biases; Richman, Kiesler, Weisband & Drascow, 1999). Anonymity and the lack of accountability that accompanies it might also reduce customers' motivation to engage in the effortful cognitive processing required to conceal or overcome any biases they may have that are based on race or gender. In contrast, supervisors who complete appraisal ratings are known to the ratee as well as the organization, their ratings are part of the employee record and they expect that their ratings will be used for employee training, feedback, and advancement decisions (Murphy, 1991). Supervisors must also justify their ratings and as such are more likely to be motivated to engage in more effortful information processing to help them reduce the influence of racial or gender bias and appear, at least superficially, to be objective.

Not only does the anonymity of customer satisfaction judgments create a situation where raters are not motivated to reduce bias, but the instructions and items used in customer satisfaction questionnaires – to form an evaluative judgment of an individual service provider or organization – may actually strengthen the effect of such biases. Stasser and Stewart (1992) found that groups participating in a murder mystery game were more likely to produce the correct answer when the

instructions were to "find the solution" versus "make a judgment." More systematic processing of information occurs when participants held a belief that there is a "necessarily correct answer." Analogously, supervisors completing a performance appraisal are frequently reminded of the importance of rating accuracy and rate on specific behavioral items (implying they should "find the solution"); that is, they frequently use behavioral rating scales (Judge & Ferris, 1993; Kane, Bernhardin & Villanova, 1995). Customers, however, are typically asked only for their "opinions" or "attitudes" about employees or the organization (suggesting they "make a judgment") (Schneider, Ehrhart, Mayer, Saltz, & Niles-Jolly, 2005). Common customer satisfaction items like "I would recommend this organization to others" and "this organization/employee meets my expectations" do not solicit recall of specific employee or organizational behaviors and so they may provide more information about the state of mind of the rater than the actual performance of the ratee or the organization.

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. Denisi, Robbins, and Cafferty (1989) argued that behavioral diaries both aid in the recall of behavioral events and improve categorization of behaviors into their appropriate domains. Customers are neither trained on how to use these techniques nor are they expected to use them when forming satisfaction judgments. As a result, the potential effects of various techniques for reducing racial and gender bias will not be observed in customer judgments.

In sum, we have argued that customer satisfaction judgments are likely to be highly susceptible to racial and gender biases because customers are anonymous, asked to make summary judgments

rather than provide accurate recall of performance-related behaviors, and are untrained in techniques that might help them overcome non-conscious biases. Our arguments lead us to expect that, in general, employees belonging to low-status demographic groups (i.e., women, racial minorities) will receive lower customer satisfaction scores than employees belonging to high-status demographic groups. But even if we found evidence for this difference, it would not be particularly informative nor would it demonstrate the operation of bias in customer judgments because it may be that the lower customer satisfaction judgments received by members of lower status demographic groups may in fact indicate lower levels of performance relative to their higher status group counterparts. Logically, customer satisfaction judgments should be at least partly influenced by employee objective performance (Wherry & Bartlett, 1982). However, if observers tend to make more favorable inferences of competence when behaviors are performed by members of high- as opposed to low-status group members, as prior studies suggest, then a more precise theoretical prediction about the effects of biases in customer judgments can be made. Namely, we hypothesize that behaviors performed by a high-status group member are likely to be interpreted more favorably by customers than *the same behaviors* performed by someone from a low-status group. Evidence for the plausibility of this hypothesis comes from studies showing that women in leadership roles are rated lower than men in similar roles (Eagly, Makhijani & Klonsky, 1992) and that ethnic minorities and women are rewarded less than whites and males for exhibiting the same advice-giving or ingratiatory behaviors (Westphal & Stern, 2007). There is also evidence that racial minorities and women who achieve equivalent levels of performance as whites and men are judged as having less underlying ability (Biernat & Kobrynowicz, 1997; Yarkin, Town & Wallston, 1982). To examine whether the biases in judgments of competence as a function of a target person's group membership found in the studies cited above generalize to customer judgments, we tested the following hypothesis:

-11-

Hypothesis 1: The relationship between employee objective performance and customer satisfaction judgments will be less positive for employees belonging to low-status demographic groups compared to employees belonging to high-status demographic groups.

The theoretical arguments supporting hypothesis 1 are based on research showing the influence of racial and gender biases when making judgments about others. Such biases would be relevant for customer judgments that ask for evaluations about an individual service provider. But we also noted earlier that another type of judgment asks for an evaluation of the context in which their customer experience takes place (e.g., cleanliness, appearance). We propose that racial and gender biases can also influence these types of judgments even though they are not associated with a particular service provider.

Our prediction draws from the notion that the positive or negative properties of an item or person can "spill over" on to the nearby context or surrounding targets (Rozin, Millman & Nemeroff, 1986). People naturally assume that physical closeness allows two objects or individuals to absorb some of the properties of the other (Frazer, 1890/1959). This idea is central to why some people carry religious symbols or good luck charms (Mauss, 1902/1972). Individuals hope the positive properties of the objects will literally rub off on them. This spillover effect also occurs in organizational contexts. For example, customers think more highly of a product if they observe a highly attractive person of the opposite sex touching it (Argo, Dahl, & Morales, 2009), and think worse of the product if they observe it in contact with a product they dislike (Morales & Fitzsimons, 2007). Similarly, the presence of low-status employees in the workplace has been shown to damage the reputations of their coworkers (Negro, Goodman & Rao, 2008). For example, Hollywood stars who appeared in films with actors later blacklisted under the "red scare" suffered reputational damage effects (Negro et al., 2008). In short, we expect customers to judge organizational contexts employing low-status employees as being

worse because customers will tend to believe that such employees' low-status is transferred to the context.

Even if customers do not think that employee demographic-based status directly spills over onto the context, they may still believe that employee status signals the level of organizational context quality (e.g., Rynes, Heneman, & Schwab, 1980; Rynes & Miller, 1983; Spence, 1973). In line with this thinking, job applicants believe that recruiters' competence and thoroughness signals an organization's overall quality (Rynes, Bretz & Gerhart, 1991). Customers may tend to view an organizational context that employs a low-status employee as signaling a worse organizational context than a context employing a high-status employee.

Combining our spillover and signaling arguments, we expect the mere presence of members of lower-status groups in the organizational environment to lead to less positive customer judgments of the service environment. The following hypothesis tests this prediction:

Hypothesis 2: Individuals will report lower customer satisfaction judgments of the

environmental context when a highly visible employee in that environment belongs to a low

(i.e., women, African-American) rather than high (i.e., male, white) status demographic group.

Not every customer will be so susceptible to racial or gender bias that they would evaluate an employee and the environment more negatively simply because they encounter an employee who belongs to a low-status demographic group. We expect the customers who are most prone to making these types of judgments to be those who hold more negative attitudes toward females or racial minorities. Our prediction is supported by the finding that people who were most effective at regulating expression of racial bias were those who had high internal motivation to do so and who also had low external motivation (Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002). Devine et al. explained their result by drawing from self-determination theory (Deci & Ryan, 1985) which assumes

people who have more autonomous, internalized reasons for regulating prejudice are most likely to do so. As we have argued previously, customers are seldom provided with external reasons for regulating racial and gender biases. Consequently, if Devine et al.'s (2002) arguments are correct then the most effective mechanisms for de-biasing customer judgments are internal. We propose that one such internal mechanism is the person's attitude toward females and racial minorities. We tested this possibility by examining whether the degree to which individuals have non-conscious, negative attitudes toward members of low-status groups (Greenwald & Banaji, 1995) would moderate the effect proposed in our first two hypotheses as stated in the following hypothesis:

> *Hypothesis 3: Low-status employees (i.e., women, African-Americans) and contexts employing such low-status employees will receive even lower customer satisfaction judgments than high-status employees (i.e., men, whites) and contexts employing such high-status employees for the same level of employee performance, when the judges have negative attitudes toward members of low-status groups.*

Thus far we have argued that customer satisfaction judgments can be influenced by recognizing a single employee's demographic characteristics. But in many cases customers interact with a variety of employees in a given customer service encounter. For example, even when doing something as simple as buying groceries, customers observe and interact with deli workers, produce employees, cashiers, and baggers. Employees in each of these positions likely have different demographic characteristics. For this reason, we suggest that another way to examine the possible influence of bias in the customer judgment process is to see whether the demographic composition of *the organizational unit* might influence customer judgments of that unit. Extending our theoretical argument to the organizational level, we expect that the degree to which an organizational unit's employees are members of low-status demographic groups will influence customer satisfaction judgments of that

organizational unit such that it will be judged as being of worse quality compared to an organizational unit whose employees are mostly from high-status demographic groups. But as we noted when predicting the influence of rater bias on judgments of individual service providers, such a finding might reflect true differences in unit performance. Hence, we make the following, more precise prediction that parallels hypothesis 1:

> *Hypothesis 4: The relationship between an organizational unit's objective performance and customer satisfaction will be less positive for organizational units that employ higher 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).*

Our theoretical arguments and hypotheses can be summarized by the conceptual model shown in Figure 1.

Insert Figure 1 here

The model shows that three attributes of the customer rating task – anonymity, the standard for making the evaluation (e.g., judgment vs. behavioral report), and the lack of training – may influence customers' vulnerability to racial and gender biases associated with membership in low- versus high-status groups. In turn, these biases can influence judgments of individual service providers. The model also shows that these judgments should partly reflect actual performance, but that racial and gender bias will moderate this relationship (hypothesis 1). Judgments of individual service providers are then shown to influence judgments of context (hypothesis 2) and, furthermore, that customer judgments of employees and the context will be moderated by the evaluator's gender and racial attitudes (hypothesis

3). Finally, our model indicates that both individual service provider judgments and contextual judgments contribute to global judgments of the organizational unit. We hypothesize that this judgment is partly influenced by the unit's actual performance, but like individual service provider judgments this relationship is moderated by racial and gender biases (hypothesis 4).

METHODS

We tested the hypotheses in our model in three studies using different samples and methods. We tested hypothesis 1 by looking at customer judgments of their physician (Medcorp Sample); we tested hypotheses 1, 2 and 3 by examining customer judgments of a bookstore and one of its employees (Bookcorp Sample); finally we tested hypothesis 4 by examining customer judgments of their golf club (Golfcorp Sample). In each sample, a large number of customers rated each of the targets.

MEDCORP SAMPLE

Our first sample was drawn from all 113 primary-care physicians (i.e., family practitioners) employed by a large health maintenance organization, hereafter referred to as Medcorp (a pseudonym). Medcorp provides coverage and healthcare for about 350,000 people in the Pacific Northwest region of the United States. Within our sample, 38.4% were women, 11.5% were ethnic minorities, and all had a medical degree. The 2006 Diversity Report by the Association of American Medical Colleges reports that 24.5% of practicing physicians are women and 12.1% are nonwhites.[2]

Measures

Medcorp routinely collected patient satisfaction ratings as well as objective behavioral indicators of physician performance that were assumed to have a direct, positive impact on patient health and well-being. This feature of our data represents a methodological improvement over studies

---

[2] However, the Association of American Medical Colleges also reports that these numbers are changing dramatically as 44% of American medical school graduates in 2006 were women and 34% were nonwhite.

that only measure employee performance with a single subjective rating and are therefore unable to determine if the rating is biased (Rotundo & Sackett, 1999). The dependent variable in our study was patient satisfaction with their physician. The independent variables were physician demographics (race and gender) and three types of patient-centered behaviors.

Customer satisfaction. A postcard survey was mailed to a percentage of each physician's patients, following doctor visits, selecting the patients so as to avoid a bias toward those patients with frequent appointments. Patients completed and returned a total of 12,091 surveys for a response rate of 52%, so that each physician was rated by an average of 107 patients. Each patient rated only one physician, so the individual ratings were independent. 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. Instead, they provided us with data indicating what percentage of each physician's patients rated the physician as "excellent." Thus, the range on this measure for each physician was from 0 to 100 percent. This measure was collected in the same quarter as all other variables.

Physician race. Medcorp identified each physician's race, and we coded whites "0" and ethnic minorites as "1." Of the 113 physicians in the sample, 10 were Asian or Pacific Islanders, two were black, and one was Native American. The percentage of ethnic minority physicians in our sample is consistent with the national average of 12.1%.

Physician gender. We coded males "0" and females "1." Forty-three of the physicians were female, which is slightly higher than the national average of 24.5%.

Objectively measured employee performance. With the growing prevalence of health maintenance organizations (HMOs) and the increasing corporatization of medicine (Feinglass & Salmon, 1990)**,** patients are increasingly being viewed by organizational administrators and physicians as customers. Therefore, physicians are increasingly being rewarded for engaging in behaviors that benefit their patients and the organization's customers (Laine & Davidoff, 1996; Stewart et al., 2000)**.** We used the customer-benefitting behaviors identified by Medcorp as our indicators of objective physician performance.

Medcorp measures customer-benefitting behaviors along three dimensions. The first is physician productivity, which is the number of health procedures performed and issues discussed in a given time period. The second is the physician's accessibility to customers measured by the number of secure emails that doctors send to customers. The third is the physician's level of quality measured by the standardized prescription rates of particular medications for customers that possess precise disease criteria. All three dimensions reflect behaviors that benefit customers by reducing the amount of time and money customers spend receiving medical care. For all metrics, physicians are shown how they compare to both the organizational goal and the organizational average. More productive physicians are able to treat more customer problems per visit, thereby saving customers' time and trips to the doctor. More accessible physicians provide greater convenience to customers who can simply email any medical questions to their physician. Higher quality physicians are better at preventing costly and deadly health events such as strokes and heart attacks. Physician compensation is tied to each of these customer-benefitting behaviors. Physicians who exceed the 40th percentile are given a bonus, while those below the 40th percentile are not given a bonus.

Physician productivity. The average number of patients seen, medical issues discussed, and medical procedures performed by each doctor in a standardized 8-hour day was recorded by the

organization's scheduling software. Medcorp physicians have a great deal of control over the amount of work that they do in a day as they can control the intensity of each visit (e.g., the number of procedures performed and patient health issues addressed per visit). The number of patients that physicians see each day is controlled by organizational administrators.

The objective performance assessment we used was the composite of average face-to-face visits and phone visits adjusted by the average intensity of each visit. Intensity was measured by Relative Value Units (RVUs), which are coded by physicians at the end of each visit according to national coding guidelines. RVUs capture the amount of time involved, the required physical and mental effort, the required judgment and technical skill, and the psychological stress entailed (Hsaio, Braun, Becker, & Thomas, 1988; Hsaio, Braun, Dunn, & Becker, 1988). Physicians check one of three RVU boxes after seeing each patient. If the patient appointment is a quick check-back or follow-up appointment, physicians check the first box, which is worth .5 RVUs. If the patient appointment involves at least two patient issues or concerns, but less than four, then the physician checks the middle box, indicating 1.0 RVUs for that visit. If the patient appointment involves five or more patient issues, then the physician checks the third box, which indicates 1.5 RVUs. According to quarterly audits by administrators, Medcorp physicians accurately record RVUs in 90 percent of patient visits. Coding errors resulting from physicians coding too many or too few RVUs are normally and equally distributed. The raw measure of productivity was standardized based on the full-time status of the physician and then multiplied by each physician's average visit intensity to obtain the quarterly *average RVU-adjusted patient encounters per day*.

Physician accessibility. The average daily number of secure emails that physicians sent to patients for the quarter was used to measure another type of customer-benefitting behavior. Patients highly value the ability to easily contact their physician. Indeed, a Harris poll has shown that 90% of

Americans who are online want the ability to e-mail their physicians, and 37% are even willing to pay for it (Taylor, 2002). Medcorp patients and physicians can communicate electronically regarding health related issues through a secure Internet health portal designed exclusively for patient-doctor communication. To use the system, patients log in to a secure website that provides the patients access to their personal health records, their lab results, and a host of health related information. Patients can send unlimited secure emails through the portal to their physician at no cost and Medcorp physicians are expected to reply to each secure email within 24 hours. Patients are encouraged to contact their physician via the system to ask basic health-related questions, to request prescription refills and to schedule follow-up appointments. Medcorp administrators think that the system saves patients doctor visits, thereby saving patients time and money. In general, physicians do not think that email improves the quality of patient care, but rather that email increases convenience for patients (Kleiner, Akers, Burke & Werner, 2002).

The Medcorp computer server automatically recorded the number of emails that each medical professional sent to his or her patients. Medical professionals had a great deal of control over how many emails they sent for two reasons: (a) they could choose which patients to make themselves accessible to, and (b) they could choose whether or not to personally respond to their patients' emails. We calculated the number of emails physicians sent per day, taking into account the number of full working days that physicians were in clinic during the time-period of this study. To enhance the normality of the variable, we used an inverse transformation and then reflected these values such that higher values represented greater use (Tabachnick & Fidell, 2003).

Physician quality. Every Medcorp primary care physician is responsible for a panel of member-patients. Of the thousands of possible treatments, prescriptions, and procedures that physicians can perform to benefit patients, one of the most important is each physician's prescription

rate of statins and angiotensin-converting-enzyme (ACE) inhibitors to patients with cardiovascular disease. Treatment of cardiovascular events such as strokes, clots, and heart attacks is the biggest healthcare cost for patients in the U.S. (Willerson & Cohn, 2000), and these drugs prevent cardiovascular events over patients' lifetimes (Gerstein et al., 2000).

According to Medcorp guidelines, all patients with cardiovascular disease should be regularly taking ACE inhibitors and some form of a statin. ACE inhibitors lower blood pressure, and statins lower cholesterol. Nationally, only 50% of all cardiovascular disease patients that should be treated with statins and ACE inhibitors are currently taking such medication (Dubois et al., 2002). These drugs significantly lower the immediate risk of a cardiovascular event (e.g., stroke, heart attack) for all individuals, regardless of sex or previous history of cardiovascular disease (LaRosa, He & Vupputuri, 1999; Yusuf, Sleight, Pogue, Bosch, Davies & Dagenais, 2000). To promote a higher prescription rate, Medcorp administrators send emails to physicians reminding them to prescribe such treatment. While prescription of these medications benefits patients by helping patients avoid death and reduce healthcare expenses, physicians often forget to prescribe these medications (Isles, 2002).

This quality variable is the composite of the percent of cardiovascular disease patients 18 years and older who were dispensed the equivalent of a 90-day supply for ACE inhibitors and statins at any time within the quarterly reporting period. The component variables approached normality, were standardized, and were added together. The resulting variable was each physician's overall *prescription rate of statins and ACE inhibitors for cardiovascular disease patients*. The average prescription rate at Medcorp is 50%, similar to the national average.

Control Variables

We controlled for several variables that were not of direct interest for testing our hypotheses but that could be theoretically related to the dependent variable and might provide plausible alternative explanations for our findings.

Average practice busyness. Patients who have to wait long periods of time to see their physician may be less satisfied, so we controlled for the busyness of each physician's practice. At the close of business each day, the Medcorp computer counts how many days into the future each physician's third available appointment is. According to the National Quality Measures Clearinghouse, counting the days until the third next available appointment is the healthcare industry's standard measure of access to care and indicates how long a patient waits to be seen. Doctors who are not very busy typically have three available appointments the next day, whereas busy doctors often do not have three available appointments for several days. The final variable was the quarterly average number of days until each physician's third open appointment slot.

Physician full-time status. We included the number of hours a physician worked in our model because patients may be more satisfied if their physician works more hours. Physicians ranged from working 30 to 100 percent of a full-time position.

Number of patients in panel. Medcorp assigns physicians to care for a particular group (i.e. panel) of patients. Patients in larger panels may be less satisfied and so we controlled for the total number of patients in each physician's panel standardized by the full-time status of the physician.

Average patient age. Older patients may have different expectations about doctor demographics, so we included the average patient age for each physician's panel in our model.

Average chronic sickness of panel. Sicker patients may be less satisfied, so we controlled for the panel chronic sickness variable calculated by Medcorp (e.g., it captures the percentage of patients with diabetes and cardiovascular disease).

Physician age and tenure. Physicians who are older or who have been employed by Medcorp for more years may have more loyal, satisfied patients.

Physician tenure by objective performance. Because women and nonwhite physicians tend to be more recently hired than male and white physicians, any influence of physician race and gender on customer satisfaction may be masked by physician tenure. We therefore included the interactions of tenure by objective performance in our models so that we could more clearly determine the interactive influence of physician race by objective performance and physician gender by objective performance on customer satisfaction.

Results

Table 1 reports the means, standard deviations, and correlation coefficients between the dependent, independent, and control variables. Our first hypothesis stated that the relationship between employee objective performance and customer satisfaction judgments would be less positive for employees belonging to low-status demographic groups compared to employees belonging to high-status demographic groups. To test this, we examined the interactions of the objective measures of employee performance (i.e. quality, productivity and responsiveness) by employee race and gender. We used hierarchical moderated regression models to do so (Aiken & West, 1991). We centered all variables involved in the interaction terms to minimize multicollinearity between the interaction terms and their individual components (Aiken & West, 1991). We entered all of the control variables in Model 1. In Model 2 we entered the control variables plus the interactions involving physician gender. In Model 3 we entered all the control variables as well as the interactions involving physician race.

Finally, in Model 4, we entered all control variables and all interaction effects. Table 2 presents the results of this analysis.

Insert Tables 1 and 2 about here

The two-way gender X objective performance interactions as a set explained a significant amount of incremental variance in the dependent variable ($R^2 = .07$, $p < .01$) providing preliminary support for hypothesis 1. Inspection of the individual regression weights showed that the physician accessibility X gender and physician quality X gender interactions were significant ($p < .05$). We probed the pattern of the interaction by examining the simple slope of the objective performance measures for male and female physicians (Aiken & West, 1991). The results of this analysis are shown graphically in Figure 2.

Insert Figure 2 about here

The figure shows a stronger positive relationship between physician customer-centered behaviors and performance ratings for men than for women. We calculated the significance of the simple slopes for interactions (Aiken & West, 1991). The coefficient of the simple slope of quality behaviors on customer satisfaction was significantly more positive for male physicians ($\underline{b} = .32$, $p < .01$) than female physicians ($\underline{b} = -.01$, n.s.). Similarly, the coefficient of the simple slope for accessibility behaviors was significantly more positive for male physicians ($\underline{b} = .13$, n.s.) than female physicians ($\underline{b} = -.17$, n.s.). Although neither simple slope for accessibility behaviors is significantly different from zero, the simple slopes are significantly different from each other ($p < .05$). By looking at the plots, one can see that the interaction is a cross-over, which shows that the direction of the relationship is the opposite, and statistically significant, for members of high- versus low-status demographic groups.

In addition, the two-way race X objective performance interactions as a set explained a significant amount of incremental variance in the dependent variable ($R^2 = .08$, $p < .05$) providing preliminary support for hypothesis 1. Inspection of the individual regression weights showed that the physician productivity X race and physician quality X race interactions were significant ($p < .05$). The forms of the interactions are shown graphically in Figure 2. Simple slope analysis reveals that the coefficient of the simple slope of quality on customer satisfaction is significantly more positive for white physicians ($\underline{b} = .29$, $p < .01$) than for nonwhite physicians ($\underline{b} = -.14$, n.s.). Likewise, the simple slope of productivity behaviors on customer satisfaction is significantly more positive for white physicians ($\underline{b} = .15$, n.s.) than nonwhite physicians ($\underline{b} = -.32$, $p < .01$). Overall, we find support for four of the relationships predicted in hypothesis 1.

Discussion

Our first sample explores whether customers, who in this case were patients of an HMO, express their race- and gender-based biases in customer satisfaction judgments. We found that objectively-measured behaviors were only positively related to customer satisfaction for physicians who were white or male. We also found that one type of customer-centered behavior was significantly *negatively* related to customer satisfaction for women and nonwhite physicians. This second finding was an even stronger result than we anticipated because logically we might expect the relationship between customer-benefiting behaviors and customer satisfaction to be weaker, but still positive, for women and nonwhites compared to men and whites.

We suggest that the observed pattern of relationships indicates that biases against nonwhite and female employees creep into satisfaction judgments. However, we must also consider this study's shortcomings. Although our Medcorp sample included a large percentage of women, it only included a small percentage of nonwhites. Moreover, many of the nonwhites were Asians rather than African-

Americans. Biases against African-Americans are more negative than those associated with Asians (Song, 2004) and so a study that included African-Americans might be better able to detect the influence of such biases on customer satisfaction judgments. We were also not able to control for employee accents or differences in employee language and communication styles, or whether customers felt certain employees had nonwhite sounding names. It is possible that the biases we observed were due to perhaps some contextual variable such as employee language skill and were not due to pre-existing customer prejudices. Finally, we did not measure whether customer raters had preexisting bias against women and minorities. That is, we had no assessment of the raters' stereotypes or racial/gender biases as potential causes of their ratings. We designed our second sample to address the limitations of our Medcorp sample. Rather than physicians, we used a type of occupation that our raters were probably more familiar with. The occupation was service employees working in a university bookstore and our raters were college students. We also used an experimental design to control for extraneous variables that might have influenced the results of our Medcorp sample.

BOOKCORP SAMPLE

In our first sample (Medcorp), we correlated objective measures of employee behavior to patient satisfaction judgments to examine the possibility that racial and gender bias might influence the latter. In our second sample (Bookcorp), student raters were asked to observe a video of an employee-customer interaction in a university bookstore, to evaluate the employee's behavior, and to provide satisfaction judgments of the store environment. Therefore our Bookcorp study is different from our Medcorp study in a variety of ways. First, we controlled for the job-related behavior (with a pre-scripted interaction) of the employee and varied only whether the behavior was performed by a male versus a female or a white versus an African-American employee. This aspect of the Bookcorp sample's design allowed us to reduce variability in employee behavior thereby providing a better test

of whether the same behavior would nevertheless produce different customer satisfaction judgments depending on the employee's gender or race. Second, we assessed how student participants, who were asked to assume the role of customers, not only evaluated the employee (as in our Medcorp sample) but also evaluated the organizational context (i.e., the bookstore) in which the employee-customer interaction took place. Third, we assessed each participant's implicit bias towards women and nonwhites to see if these implicit attitudes were causes of gender or racial bias in the ratings.

Sample

Eighty-six university students from a major northwestern U.S. public university watched two videos of a university bookstore employee interacting with a customer and were asked to evaluate the employee and the bookstore. The bookstore in the video clips was from a large East Coast U.S. university and it is highly unlikely that any of the participants had visited the bookstore previously. The "employees" and "customers" were hired professional actors and the scripted interaction was filmed before the bookstore opened in the morning (although our raters taking the customer perspective were not aware of this). We assigned 33 participants to view the white male, 21 to view the white female employee, and 34 to view the black male employee. Overall, a substantial percentage of our participants were nonwhite (43 percent) or female (38 percent). By having a heterogeneous sample of raters, the subjects are perhaps more representative of the population of people using the book store (as compared to all white males).

Design

Our design was a mixed-factorial design with one between subjects factor (Employee Demographic Characteristics) and one within subjects factor (Employee-Customer Interaction). We treated Employee Demographic Characteristics as a between-subjects factor to reduce participant awareness that they were participating in a race or gender related study. We presented all participants

with two videos depicting different employee-customer interactions. One video involved the employee ringing up a book and telling the customer that the book's price in the computer was higher than its price on the shelf. The other video involved the same employee trying to help a customer find a book the customer wanted. Each video was about one minute in length. Each participant saw both videos of the same employee. We randomly assigned the ordering of the videos within each condition and found no evidence of an order effect. The customer and employee interaction was from a written script to ensure that their behavior was equivalent across conditions. The store background was also held constant across conditions since the camera was in the same location when filming the different employees for each interaction.

Dependent Variables

Customer satisfaction with the employee. Our measure of customer satisfaction with the employee asked raters to identify on a 7-point Likert scale (1 = very poor; 7 = excellent) how satisfied they were with (1) speed of service; (2) quality of service; (3) availability of staff for assistance, and; (4) employee responsiveness to customers' issues and concerns. This measure was adapted from an existing customer satisfaction survey we obtained from a large organization (see Appendix 1 for customer satisfaction items used across the three studies). Coefficient alpha for this measure was .74.

Customer satisfaction with the context. Our measure of customer satisfaction with the context asked raters to identify on a 7-point Likert scale (1 = very poor, less than expected, definitely would not, or strongly disagree; 7 = excellent, better than expected, definitely would, or strongly agree) how satisfied they were with (1) the bookstore's appearance; (2) the degree to which the bookstore was conducive to learning; (3) whether the bookstore had up to date equipment; (4) the degree to which the bookstore's facilities were visually appealing; (5) whether the bookstore's appearance was in keeping with the type of services provided; (6) the bookstore relative to their expectations, and; (7) their

likelihood of recommending the bookstore to others. This measure was also adapted from an existing customer satisfaction survey we obtained from a large organization. Coefficient alpha for this measure was .76.

Predictor Variables

Condition. We had two conditions—one for race and one for gender. The gender condition included participants who viewed the white male employee or white female employee (1 = participants viewed two videos of a white woman employee, 0 = participants viewed two videos of a white man employee). The race condition included participants who viewed either the white male employee or the nonwhite male employee (1 = participants viewed two videos of a nonwhite man employee; 0 = participants viewed two videos of a white man employee).

Implicit bias. To measure raters' racial and gender prejudices, we administered two Implicit Attitudes Tests (IATs). The IATs were constructed to capture each participant's level of non-conscious bias against nonwhites and women (Greenwald, Nosek and Banaji, 2003). We chose to use the IAT as opposed to other types of bias measures (e.g., modern racism scale) because it is more difficult for participants to hide prejudices on the IAT than on explicit measures (Nosek, 2005). The gender IAT was administered after the participants saw the videos and made their customer satisfaction judgments, but the race IAT was administered between the videos and the ratings. There is no evidence of order effects for the IAT and dependent variables—probably because subjects still respond in socially desirable ways on the explicit measures (Greenwald, Poehlman, Uhlmann & Banaji, 2009).

Importantly, implicit attitudes appear to be better predictors of behavior than their explicit counterparts, especially when social sensitivity concerns are high (Greenwald et al., 2009). For instance, implicit (but not explicit) attitudes about African Americans have been shown to predict desire to work with an African American partner on an intellectual task (Ashburn-Nardo, Knowles, &

Monteith, 2003), and nonverbal actions (eye contact and other "friendly" behaviors) toward African American interaction partners (McConnell & Leibold, 2001). Though the correlation between implicit and explicit attitudes varies across domains (Nosek, 2005), the predictive validity of each suggests that they represent independent processes that explain unique variance in behavioral outcomes (see Greenwald et al, 2009, for a meta-analysis of the predictive validity of the IAT).

Control Variables

We controlled for rater race, gender and age to account for rater demographics which might plausibly influence reactions to employee demographics.

Results

We regressed the customer judgments of the employee and the organizational context on our controls, predictors, and interaction to determine the degree to which customer judgments of the employee and the organizational context reflected race and gender bias. Tables 3 and 4 present the regression models we used to test hypotheses 1, 2 and 3.

Insert Tables 3 and 4 about here

Hypothesis 1 states that the relationship between employee objective performance and customer satisfaction judgments will be less positive for employees belonging to low-status demographic groups compared to employees belonging to high-status demographic groups. Because objective performance was held constant due to the employee script, the main effects of employee race and gender on customer satisfaction with the employee were used to test this first hypothesis. Model 2 in Table 3 shows that raters taking the customer perspective were significantly less satisfied with women employees than their equally well-performing white male counterparts ($\Delta R^2 = .06$; $\underline{b} = -.28$; $p <$

.05). However, we did not find evidence of bias in customer satisfaction judgments of the nonwhite employee ($\underline{b}$ = -.02; n.s.). Overall, we found some support for hypothesis 1.

Hypothesis 2 states that people would report lower customer satisfaction judgments of the store environment when an employee in that environment belongs to a low (i.e., female, African-American) rather than high (i.e., male, white) status demographic group. Model 2 of Table 4 shows there is a significant main effect of race and gender on judgments of the store environment. Indeed, Model 2 of Table 4 shows a main effect of the female condition (ΔR^2 = .17; $\underline{b}$ = -.45; $p < .01$) and the nonwhite condition (ΔR^2 = .15; $\underline{b}$ = -.44; $p < .001$), suggesting that raters' biases influence judgments of the organizational context. We found strong support for hypothesis 2.

Hypothesis 3 suggests that people would report even lower customer satisfaction judgments of the employee and the store environment when observing an employee belonging to a low-status demographic group when the customer has negative implicit attitudes toward that group. Model 3 of Table 3 shows that the coefficient for the interaction term IAT Score X Nonwhite Condition is significant and in the expected direction for customer satisfaction with the employee (ΔR^2 = .08; $\underline{b}$ = -.28; $p < .01$). To gain more insight into the nature of this effect we plotted the interaction and analyzed the simple slopes (see Figure 3). Individuals with high levels of implicit bias (+1 s.d.) were significantly more likely to report lower satisfaction with the nonwhite male's performance than with the white male's ($p < .01$). However, the coefficient for the interaction term IAT Score X Gender Condition was not significant for customer satisfaction with the employee.

As for customer satisfaction with the context, Model 3 of Table 4 shows that the coefficient for the interaction term IAT Score X Race is significant and in the expected direction for customer satisfaction with the context (ΔR^2 = .04; $\underline{b}$ = -.18; $p < .05$). We plotted the interactions and conducted a simple slope analysis (see Figure 3). Customer IAT score (+1 s.d.) was positively related to customer

satisfaction with the context when customers were observing a white male employee ($\underline{b} = .33$; $p < .01$) but was negatively related to customer satisfaction with the context when customers were observing a nonwhite male employee ($\underline{b} = -.21$; $p < .05$). The coefficient for the interaction term IAT Score X Gender Condition was significant and in the expected direction for customer satisfaction with the context ($\Delta R^2 = .04$; $\underline{b} = -.23$; $p < .05$). Customer IAT score (+1 s.d.) was positively related to customer satisfaction with the context when customers were observing a white male employee ($\underline{b} = .23$; $p < .05$) but was negatively related to customer satisfaction with the context when customers were observing a white female employee ($\underline{b} = -.21$; $p < .05$). Overall, we found three significant coefficients supporting hypothesis 3. These results and plots suggest that judgments of employees and judgments of the organizational context are vulnerable to non-conscious biases.

Discussion

We found that raters taking the customer perspective rated the employee and the organizational context as being worse when observing the performance of a low-status employee and this was especially true if the person held implicit biases about that low-status group. We are encouraged that the results regarding hypothesis 1 were not as strong as we anticipated because it gives us hope that extinguishing these biases may be straightforward. Specifically, the laboratory context was less anonymous than a typical customer satisfaction questionnaire setting, which may have weakened the influence of bias on customer satisfaction judgments of employees. Although we told the participants that their responses were anonymous, they may have felt scrutinized because they provided their judgments when an experimenter was present, and wrote their names on a separate sign-in sheet to receive participation credit for a class. We observed that the effects of bias on judgments of the organizational context were still quite strong, which makes sense to us because participants may have

been unaware of and therefore unable to suppress their biases which spilled over onto the organizational context.

The results from our Bookcorp sample are consistent with ideas suggesting that contemporary forms of racial and gender bias may assume covert forms. Logically, one might expect that judgments of the organizational context, which are very distal from low-status employees, would be least influenced by rater biases. However, we believe that some customers may think that the negative properties of low-status employees spill over onto the organizational context and also that the presence of low-status employees may signal to customers that the organization as a whole is low-status.

As summarized by hypothesis 4, we expected that the relationship between an organizational unit's objective performance and customer satisfaction would be less positive for organizational units that employ higher percentages of employees belonging to low-status demographic groups (i.e., women and racial minorities) compared to units that employ higher percentages of employees belonging to high-status demographic groups (i.e., men and whites). By returning to the field to test this hypothesis, we complete the full-cycle of research model and assess the generalizability of our theory to a different organization.

GOLFCORP SAMPLE

Our sample was drawn from a large country club organization, hereafter referred to as Golfcorp.[3] Golfcorp has 66 country clubs across the United States and roughly 70,000 customer-members. It directly employs approximately 8,000 employees to care for its customer-members. Our sample consisted of all 66 Golfcorp country clubs. Within our sample, 31.4% of employees were women, 18.1% were Latino, 6.7% were African-American, and 1.7% were Asian-American or Native-American.

Measures

---

[3] Golfcorp is a pseudonym.

Golfcorp routinely collects customer satisfaction ratings as well as objective indicators of facility performance that were assumed to have a direct, positive impact on customers' service experiences. The dependent variable in our study was customer satisfaction with the facility. The independent variables were each club's employee demographics (percentage nonwhite and female employees) and two types of objective club performance.

Customer satisfaction with facility. Like many organizations, Golfcorp measures customer satisfaction with a quarterly survey, which is mailed to a percentage of each facility's customers. An average of 63.8 customers rated each facility. The average response rate per facility was 27.3 percent (an average of 234 surveyed customers per facility). The marketing company hired to do the customer survey randomly sampled each facility's customers each quarter until they got either 20 respondents or three percent of the total customer base (whichever was larger). The items used for this measure reflect a focus on the facility context (quality of its clubhouse and golf course) and overall ratings of the facility, similar to what was used in the Bookcorp sample. Customers rated each of the items on a 5-point Likert scale (1 = very poor; 5 = very good) "How would you rate the following aspects of your club…" (1) Maintenance of grounds/Appearance of clubhouse; (2) Locker rooms and restrooms; (3) Quality of greens; (4) Condition of course; (5) Pace of play; (6) Condition of practice facilities; (7) Ability to obtain desired tee times; (8) Club meets expectations (1 = less than expected, 5 = better than expected), and; (9) Likelihood of recommending club to others (1 = definitely will not, 5 = definitely will). Coefficient alpha for this performance measure is .81. In our analyses, we lagged this measure six months after the independent variables to more conclusively show that employee demographics and objectively-measured performance influence customer ratings, rather than the other way around.

Employee race. Golfcorp identified the percentage of white and nonwhite employees in each facility. Across the 66 clubs in the sample, 26.5 percent of employees were nonwhite. According to

the U.S. Census Bureau's 2000 census, the percentage of ethnic minority employees in our sample is consistent with the national average of 28 percent of the population.

Employee gender. We also obtained this variable from Golfcorp records. Thirty-one percent of employees in our sample were women, which Golfcorp leaders believe to be consistent with the country club industry average. However, the percentage of women in our sample is lower than the percentage of women across all industries, which is 46%, according to the Bureau of Labor Statistics at the U.S. Department of Labor.

Objective facility characteristics. As in our first two samples, we wanted to clearly identify the portion of variance attributable to customer bias versus the portion attributable to better facility performance. We therefore used two attributes as our indicators of objective facility performance— facility productivity and facility attribute quality. Both dimensions reflect facility characteristics that benefit customers. Facilities with more productive employees create more value—both for Golfcorp and for customers. Indeed, Golfcorp executives told us that facilities with higher productivity values charge lower dues to members, are more profitable and are simply better-run facilities. Higher attribute quality benefits customers by allowing customers to enjoy newer and better facilities. Facility managers are shown how they compare to other facilities in terms of quality and productivity. Employee compensation is tied to the productivity measure, but not the quality measure. Employees in facilities that are above average in productivity are given a bonus, while employees at below average clubs are not given a bonus.

Facility Productivity. Facility productivity was calculated by Golfcorp's central accounting office for the calendar year ending six months before the dependent variable was collected. This variable is each club's annual profits divided by the average number of employees working for the club in that year. The number of employees at each club is centrally controlled such that clubs with more

members are allotted proportionally more employees by the central office. Therefore facility

productivity is determined by the employees' effectiveness at creating value.

Facility quality attributes. Over time, the condition of the golf course and the clubhouse

deteriorates and needs to be rebuilt or refurbished. Golfcorp assesses the quality of the course and

clubhouse of each club to ensure that customers are receiving a high standard of service. Golf courses

are assigned a quality rating based on the percentage of the course that is infiltrated by crab grass and

dead spots (1 = more than 40% of course is crab grass or dead spots; 5 = less than 5% of course is crab

grass or dead spots). Likewise, clubhouses are assigned a quality rating, indicating how recently they

were built or refurbished (1=built or refurbished more than 15 years ago; 5 = built or refurbished 2

years ago or less). The overall facility attribute quality variable is the composite of the clubhouse

quality rating and the golf course quality rating. The two component variables approached normality

and were added together. The resulting variable was each facility's overall *golf course and clubhouse*

quality.

Control Variables

We again controlled for several variables that might provide plausible alternative explanations

for our findings.

Facility size. Larger facilities may have less satisfied customers because the atmosphere may

be less personal, so we controlled for the number of employees in our analysis.

Average employee age. Customers may favor younger employees, so we controlled for

employee average age in the analysis.[4]

Percentage of temporary employees. Permanent employees, as opposed to temporary ones,

may have a better idea of how to get things done around the facility and thus be more satisfying to

---

[4] Although in our Medcorp sample we suggested that customers might prefer older physicians because they have greater experience, we expect that country club customers might prefer younger country club employees because such employees might be more eager to please customers.

customers. We controlled for the working status of each employee (0=temporary, 1=permanent) at the time the objective facility performance variables were measured.

Average customer tenure. Longer-tenured customers may be more familiar with the club and therefore be more satisfied, so we controlled for the average number of months the customers indicated they had been members of the organization.

Percentage of male customers. Because these facilities were golf clubs, male members may be more satisfied than female members. We therefore controlled for the percentage of customers who indicated they were male on the customer satisfaction survey. Although we would have liked to control for customer race, the organization does not collect that information.

Average customer age. Finally, older customers may be wealthier and generally more satisfied than younger ones so we controlled for the average customer-reported age.

Results

Table 5 reports the means, standard deviations, and correlation coefficients between the dependent, independent, and control variables. We used hierarchical moderated regression models to examine the hypothesized interaction effects (Aiken & West, 1991). We centered all variables involved in the interaction terms to minimize multicollinearity between the interaction terms and their individual components (Aiken & West, 1991). Table 6 presents the results of our hierarchical moderated regression analysis. We entered all of the control variables in Model 1. In Model 2 we entered the interactions involving gender and the two dimensions of objective performance, In Model 3 we entered the interactions involving race and the two dimensions of objective performance, and in Model 4 we entered all four interactions.

Insert Tables 5 and 6 about here

Hypothesis 4 stated that the association between an organizational unit's objective performance and customer satisfaction would be less positive for organizational units that employ higher percentages of employees belonging to low-status demographic groups compared to units that employ higher percentages of employees belonging to high-status demographic groups. The two-way gender X objective performance interactions as a set explained a significant amount of incremental variance in the dependent variable ($\Delta R^2 = .04$, $p < .05$) providing some further support for this hypothesis. Inspection of the individual regression weights in the full model showed that the facility quality X gender and facility productivity X gender interactions were significant ($p < .05$). We probed the pattern of the interactions by examining the simple slopes of the objective performance measures for facilities with high and low percentages of female employees (Aiken & West, 1991). The results of this analysis are shown graphically in Figure 4.

Insert Figure 4 about here

Figure 4 shows a stronger positive relationship between objective performance and customer satisfaction for facilities that have a low percentage of female employees than for facilities that have a high percentage of female employees. Facility quality was significantly more positively associated with customer satisfaction for facilities with a low percentage of female employees ($\underline{b} = .13$, n.s.) than for facilities with a high percentage of female employees ($\underline{b} = -.14$, n.s.). Likewise, facility productivity was more positively associated with customer satisfaction for facilities with a low percentage of nonwhite employees ($\underline{b} = .11$, n.s.) than for facilities with a high percentage of nonwhite employees ($\underline{b} = -.11$, n.s.). Again, even though these simple slopes are not significantly different from zero, the significant regression coefficient in the full model demonstrates that they are significantly different from each other. By looking at the plots, one can see that the interaction is a cross-over,

which shows that the direction of the relationship is the opposite, and statistically significant, for members of high- versus low-status demographic groups. These results support hypothesis 4 for both of the gender X objective performance relationships.

The two-way race X objective performance interactions as a set explained a significant amount of incremental variance in the dependent variable ($\Delta R^2 = .15$, $p < .05$) providing preliminary support for our hypothesis. Inspection of the individual regression weights from the full model showed that the facility attribute quality X race and facility productivity X race interactions were significant ($p < .05$). Simple slope analysis revealed that the association between facility quality and customer satisfaction was significantly more positive for facilities employing a low percentage of nonwhites ($\underline{b}= .15$, n.s.) than for facilities employing a high percentage of nonwhites ($\underline{b}= -.14$, n.s.). Likewise, the association between facility productivity and customer satisfaction was more positive for facilities employing a low percentage of nonwhites ($\underline{b}= .32$, $p < .01$) than for facilities employing a high percentage of nonwhites ($\underline{b}=-.23$, $p < .05$). These results support hypothesis 4 for both of the race X objective performance relationships.

Discussion

In this Golfcorp sample, we found that objectively measured behaviors that benefit customers were positively related to customer satisfaction, but *only* for facilities with a low percentage of nonwhite and female employees. These results parallel the results of our Medcorp sample.

GENERAL DISCUSSION

We set out to determine if and how customer satisfaction judgments are influenced by racial and gender biases. Across three samples we found converging evidence that customer satisfaction judgments are susceptible to systematic and predictable racial and gender biases. Customers tended to be less satisfied with the services provided by women and nonwhite employees than by men and white

employees, even when controlling for objective indicators of performance (Medcorp sample). We also

found that these biases operated on judgments about the store context when a third-party evaluator

observes an employee interacting with a customer (Bookcorp sample), particularly if the observer

holds negative implicit attitudes toward women or minorities. Finally, we found that evaluations of the

organizational unit as a whole are negatively associated with the presence of nonwhite and female

employees (Golfcorp sample). It is worth noting that we found evidence for the operation of racial

biases regardless of whether the nonwhite employees were predominantly Asian (Medcorp sample),

African-American (Bookcorp sample), or Latino (Golfcorp sample).

Taking a "full-cycle" research approach, we began by asking the question of whether customer

judgments of their direct service provider were biased against women and minorities in a field sample

of physicians. We then attempted to replicate our findings in the laboratory and to extend them by

testing whether less direct judgments about the context in which the service encounter takes place are

susceptible to racial and gender bias. These ideas were confirmed with student subjects observing a

familiar retail exchange, but not directly experiencing it. Finally, we went back into the field setting to

see whether customer judgments of the organizational context and overall judgments of the unit

providing the service also contained bias. Our approach follows Cialdini's (1995) recommendation that

the best way to gain a fuller a understanding of a phenomenon is to conduct one laboratory study and

two field studies so that field-generated and experimentally-verified findings can be re-validated in a

different field setting.

The consistency of our results across three different samples and methodologies testifies to the

robustness and generality of the systematic biases we observed and to the internal validity of our

theoretical model. The pattern of these biases may help explain the persistence of demographic

inequalities in organizations. To cite just a few examples, women and nonwhites make 25 percent less

than their male and white counterparts in equivalent jobs (U.S. Census Bureau, 2006), women and

nonwhites are twice as likely as white men to be unemployed and underemployed (NIOSH, 2002), and

women and ethnic minorities are not well represented among the ranks of highly paid managers and

professionals in U.S. corporations and in prestigious occupations like law and medicine (e.g., Baldi &

McBrier, 1997; Eagly & Karau, 2002; Wilson, Sakura-Lemessy & West, 1999). Economists have often

been perturbed by these demographic inequalities because orthodox economic theory would predict

that some of these inequalities should be erased when employers compete for women and nonwhite

applicants whose wages are 25 percent less costly than their white and male counterparts (*The

Economist*, 2008). However, our results suggest that there may be valid reasons why organizational

decision makers might conclude that their customers will be less satisfied with the services provided

by nonwhite and female employees than white and male employees, and that these judgments can

influence the satisfaction with the units in which these employees reside. If so, then it makes economic

sense from the company's perspective to favor white and male employees. Evidence from highly

publicized lawsuits (e.g., Shoney's Restaurant; Abercrombie & Fitch) suggests that managers are

keenly aware of the fact that some customers may prefer white and male employees. Executives in

these cases admitted to deliberately favoring white employees in hiring and promotion decisions to

enhance customer satisfaction and organizational profitability (Brief et al., 2000).

When employees or units are viewed as performing less well by customers as a result of

demographic characteristics like race or gender, one could argue that these employees and units *should*

receive fewer rewards, bonuses, and promotional opportunities. But it is important to note that the

customer judgments in our studies were inconsistent with other objective indicators of performance. In

other words, nonwhites and women may have behaved the same way as their white and male

counterparts in trying to provide satisfactory customer service, but if compensation and other

organizational benefits are linked to customer satisfaction ratings then they may not be rewarded similarly for identical behavior, which would violate the principle of equity which most business enterprises claim to follow.

Our results suggest that if customer evaluations become widely and uncritically used to determine pay and promotion opportunities, the job outcomes of women and ethnic minorities could be adversely impacted. For example, consider what would happen if managers notice which employees routinely receive the highest customer satisfaction scores and use this information to make promotion decisions. At higher levels of the organization, executives may examine which of their organizational units achieve the highest levels of customer satisfaction and promote those managers further up the organizational hierarchy. Our data suggest that one possible consequence of these decisions is that whites and men will be much more likely than their nonwhite or women counterparts to receive favorable customer satisfaction judgments, which should accelerate their journey up the organizational ladder. Likewise, managers who tend to stock their organizational units with whites and men are likely to have more career success than managers who do not.

Our finding that customer biases can spill over onto the surrounding organizational context contributes to the literature on contamination and signaling and also illustrates the subtle operation of racial and gender bias. Marketing researchers have shown how observable customer characteristics like physical attractiveness can influence other customers' desire to purchase a product (Argo et al., 2009). To our knowledge, ours is the first study to show how an observable characteristic of an employee like race or gender can influence customer perceptions of an organization's contextual quality. This finding may provide insight into phenomenon known as "white flight" where whites move out of a neighborhood once a critical mass of nonwhites move in (Gladwell, 2000; Kruse, 2005). In an organizational setting, a similar phenomenon may operate in which customers may link

conscious or unconscious negative attitudes they have towards members of status groups to employees who belong to these groups. In turn, these associations "contaminate" customer perceptions of the organizational context. This process of contextual spillover may partly explain why managers have often been reluctant to pursue diversity despite the known performance advantages of having a diverse workforce (Joshi, Liao & Jackson, 2006; O'Reilly, Williams, & Barsade, 1997). It may be that managers are aware that diversity has hidden costs because it increases the possibility of "customer flight" to an organization that has fewer employees who belong to low-status demographic groups. Our data do not allow us to draw this conclusion so at this point we are merely speculating on a possible consequence of the racial and gender biases we did find. However, we believe that future research should investigate this possibility and also take into account how a *customer's* own race or gender might influence their satisfaction judgments of employees who are demographically similar or dissimilar. Future studies using customer satisfaction as an outcome variable should take into account the demographic make-up of the employees as well as objectively measured organizational characteristics.

Our findings cast doubt on the ability of customers to accurately perceive the quality of customer service organizations. The theoretical attributes we suggested as possible causes for our data were meant to explain why customer biases can occur, but like any useful theory they also suggest potential remedies for these biases. Based on our theorizing, racial and gender biases in customer satisfaction judgments may be reduced by (1) making customer ratings less anonymous; (2) changing the standards customer use to make their ratings so they emphasize behavior rather than subjective judgments; and (3) introducing customer de-biasing education or training in the evaluation process. Based on prior research regarding bias in ratings, we believe these three factors are the most important variables for reducing bias in customer satisfaction judgments. Among the three, de-biasing education

and training may seem least practical, but this might not necessarily be the case if it was done on a select group of customers who voluntarily agreed to go through such training.

In addition to addressing factors that cause bias in customer ratings, organizations can take practical steps to minimize the potential adverse impact of customer biases on nonwhite and female employees' careers. For example, organizations might consider only using satisfaction surveys from frequent customers to ensure that raters have sufficient exposure to targets. Organizations could also ask for customer feedback during the customer service encounter so that customers will be most likely to be paying attention and less likely to rely on information subject to memory bias when judging their customer experience. Organizations might also want to let customers know that the data will be used to make career progression decisions so that customers are more motivated to judge responsibly. Organizations could also insert "bias sensitive" questions in customer satisfaction judgments so that responses from potentially biased customers can be given less weight or discarded. Alternatively, organizations may be able to statistically correct for bias when calculating customer satisfaction judgments. Finally, using different survey formats for customer rating scales might also be helpful for circumventing rater biases, like forced-choice, behaviorally anchored rating scales (citing specific valued behaviors), and un-weighted and weighted checklists. Organizations should consider the tradeoffs between these formats and choose the one that is most likely to reduce the effects of customer judgment biases on the career prospects of those who are most vulnerable to being targets of such bias.

Limitations

We believe our findings provide strong, consistent support for our theoretical predictions. However, like all research, ours has its share of limitations. First, role congruence may be an issue in our Medcorp sample as patients may expect their physicians to be white and male. However, there is not much evidence to suggest that a white male is congruent with the doctor role. Indeed, patients

prefer their doctor to look like them (women prefer women doctors, and nonwhite patients prefer nonwhite doctors; Chen, Fryer, Phillips, Wilson & Pathman, 2005; Cooper-Patrick, Gallo, & Gonzales, 1999). Likewise, role congruence should not be an issue in our Bookcorp sample, so our consistent results across these two samples provide us with some confidence that role-congruence alone is not responsible for our findings. Still, future research would be well-served to test our hypotheses in a greater variety of samples. Another issue is that we did not control for patient gender and race in our Medcorp sample. Although we had these variables and including them in the analysis slightly strengthened our results, we chose to exclude them from our final analysis due to multi-collinearity problems (i.e., the respective correlations between customer gender/race and physician gender/race were greater than .90). We would have liked to include customer race as a control variable in our Golfcorp sample, but Golfcorp does not collect this variable. However, our Golfcorp results were identical regardless of whether we included customer gender or not, which gives us confidence that our results are accurate. Moreover, we found in our laboratory sample that nonwhite and women raters were no less biased in their ratings than whites or men**.**

Another potential limitation is that unobserved variables may be responsible for our results. Particularly in our Golfcorp sample, employee demographics may have masked the facility's strategy. Facility executives pursuing a low-cost strategy may have hired a large number of women and nonwhite employees, whereas those pursuing a premium-pricing strategy may have hired a large number of whites and men. We added and then struck nine control variables to test this idea: (1) average employee tenure; (2) annual employee voluntary turnover rate and (3) annual employee involuntary turnover rate; (4) whether the facility has a pool; (5) whether the facility has a fitness center; (6) whether the facility has tennis courts; (7) average employee satisfaction with their pay; (8) average employee rating of the quality of service customers receive from their facility, and; (9) percent

of customers who purchased the expensive "society" membership which gives customers privileges at other clubs and increases the quantity and quality of services customers receive from their home club each month. Specifically, we believed clubs pursuing a low-cost strategy might have higher voluntary and involuntary turnover (because of lower quality employees), lower average employee tenure, fewer amenities (i.e., tennis, fitness, pool), fewer employees who are satisfied with their pay, fewer employees thinking the facility provides a high quality of service, and fewer customers buying the expensive society membership. We ran the interactions of each of these variables by objective performance and our four race and gender interactions remained significant, but none of the nine additional main effect variables or 18 additional interactions were significant. To maintain compliance with the rule of thumb that there should be five cases per variable (Tabachnick & Fidell, 2003), we do not report any of these main effects or interactions. Relatedly, we also did not test or report whether nonwhite women face a double jeopardy in terms of customer satisfaction to maintain compliance with the five to one rule of thumb (i.e. we were not able to test the three-way interactions involving employee objective performance, employee race and employee gender).

We should also mention that our method of testing for possible bias in performance evaluations was a significant improvement over past studies. First, we used an objective performance standard so that we could compare subjective judgments to this standard and therefore determine whether the customer judgments of performance might be influenced by race and gender. Second, our subjective judgments were based on aggregated judgments from a large number of customers rather than relying on the judgments of a single supervisor. This is important because the large number of raters provides a highly reliable subjective performance rating for each individual, context, or organization. Finally, we controlled for several variables that could provide alternative explanations for our results, such as the average employee age and the average customer tenure with the organization.

Given these methodological strengths of our research, it is unsettling to find that customers may not respond favorably to organizational characteristics designed to benefit them when these organizations have a high percentage of low-status employees. At Golfcorp, employees at clubs with a high percentage of female and nonwhite employees can in fact be economically harmed by customer satisfaction evaluations because clubs who fail to achieve the target level of customer satisfaction (i.e., below the organizational average) do not receive a salary bonus. The practical implications of our results become more apparent when we when examine the effect sizes in our sample. Across our three studies, the racial and gender bias effects on customer satisfaction judgments explained between 15 and 24 percent of the variance in customer satisfaction. Cohen (1988) provides ballpark descriptors of effect sizes based on R-squared values—"large" ($R^2 = .25$), "medium" ($R^2 = .09$), and "small" ($R^2 = .01$). Therefore, the average observed effect size of racial and gender bias across our three samples is between medium and large.

Conclusions

In these different samples we demonstrated that customer ratings are biased against women and racial minorities. We had two field studies and one laboratory study utilizing a full cycle research strategy. The effects were demonstrated for three different minority groups and three different contexts involving employee-customer contact. In all three settings we controlled for actual objective behavior or performance along with a series of other controls appropriate for that context. The effects are demonstrated for individual targets as well as the context or organization in which the targets work. In the laboratory sample we showed that the biased ratings are exacerbated by implicit racial or gender bias. In short, these are fairly robust findings across jobs, contexts, raters and ratees.

If these results are replicated and generalizable they have significant implications for organizational practice. If managers are serious about the fair treatment of their employees and the

promotion of diversity they will need to treat customer ratings differently. More specifically, the

rating process can be changed by increasing information, responsibility or training for raters or by

changing how customer ratings are used. In the latter context, organizations can perhaps measure and

discount such biases or use statistical procedures to adjust the ratings to remove the bias. Without such

actions, given the increasing dependence on customer ratings, we are likely to not only maintain

existing levels of inequitable compensation and advancement for women and minorities, we are likely

to increase these inequities. This outcome is unacceptable in a society that is committed legally,

morally and socially to fair treatment for all in the workplace.

References

Aiken, L. S. & West, S. G. 1991. *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage Publications.

Anderson, E. W., Fornell, C., & Mazvancheryl, S. K. 2004. Customer satisfaction and shareholder value. *Journal of Marketing,* 68: 172-185.

Anderson, E. W., & Mittal, V. 2000. Strengthening the satisfaction-profit chain. *Journal of Service Research,* 3: 107-120.

Argo, J. J., Dahl, D. W., & Morales, A. C. 2009, Positive Consumer Contamination: Responses to Attractive Others in a Retail Context. Forthcoming at the *Journal of Marketing Research*.

Ashburn-Nardo, L., Knowles, M. L., & Monteith, M. J. 2003. Black Americans' implicit racial associations and their implications for intergroup judgment. *Social Cognition, 21*: 61-87.

Aquino, K., & Bommer, W. H. 2003. Preferential mistreatment: How victim status moderates the relationship between organizational citizenship behavior and workplace victimization. *Organization Science,* 4: 274-285.

Baldi, S., & McBrier, D. B. 1997. Do the Determinants of Promotion Differ for Blacks and Whites? Evidence from the U.S. Labor Market. *Work and Occupations,* 24: 478-497.

Baltes, B. B., & Parker, C. P. 2000. Reducing the effects of performance expectations on behavioral ratings. *Organizational Behavior and Human Decision Processes,* 82: 237-267.

Berger, J., Conner, T. L., & Fisek, M. H. 1974. *Expectation States Theory: A Theoretical Research Program.* Cambridge, MA: Winthrop.

Berger, J., Fisek, M. H., Norman, R. Z., & Zelditch, M., Jr. 1977. *Status Characteristics and Social Interaction: An Expectation States Approach.* New York: Elsevier.

Bertrand, M., & Mullaninathan, S. 2004. Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review,* 94: 991-1013.

Biernat, M., & Kobrynowicz, D. 1997. Gender and race-based standards of competence: Lower minimum standards but higher ability standards for devalued groups. *Journal of Personality and Social Psychology,* 72: 544-557.

Bobo, L. 1998. Race, interests, and beliefs about affirmative action – Unanswered questions and new directions. *American Behavioral Scientist,* 41: 985-1003.

Bracken, D. W., Church, A. H., & Timmreck, C. W. 2001. *Handbook of Multisource Feedback: The Comprehensive Resource for Designing and Implementing MSF Processes.* New York, NY: Jossey-Bass.

Brief, A. P., Dietz, J., Cohen, R. R., Pugh, S. D., & Vaslow, J. B. 2000. Just doing business: Modern racism and obedience to authority as explanations for employment discrimination. *Organizational Behavior and Human Decision Processes, 81,* 72-97.

Chatman, J. A., & Flynn, F. J. 2005. Full-Cycle Micro-Organizational Behavior Research. *Organization Science,* 16: 434-447.

Chen, F.M., Fryer, G.E., Jr., Phillips, R. L., Jr., Wilson, E., & Pathman, D.E., (2005). Patients' Beliefs About Racism, Preferences for Physician Race, and Satisfaction With Care. *Annals of Family Medicine.* 3: 138-143.

Cialdini, R. B. 1995. A full-cycle approach to social psychology. In G. G. Brannigan & M. R. Merrens (Eds.), *The Social Psychologists: Research Adventures:* 53-72. New York: McGraw-Hill.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences (2nd ed.).* Hillsdale, NJ: Lawrence Erlbaum.

Cooper-Patrick, L., Gallo, J.J., & Gonzales, J.J.,(1999). Race, gender, and partnership in the patient-physician relationship. *Journal of the American Medical Association.* 282: 583-589.

Crandall, C. S., & Eshelman, A. 2003. A Justification-Suppression model of the expression and experience of prejudice. *Psychological Bulletin,* 129: 414-446.

Davis, W. E., & Davis, D. R. 1999. The University Presidency: Do Evaluations Make a Difference? *Journal of Personnel Evaluation in Education,* 13: 119-140.

Deci, E. L. & Ryan, R. M. 1985. The general causality orientations scale – Self-determination in personality. *Journal of Research in Personality,* 19:109-134.

DeNisi, A. S., Robbins, T., & Cafferty, T.P. 1989. Organization of information used for performance appraisals: Role of diary keeping. *Journal of Applied Psychology,* 74: 124-129.

Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. 2002. The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. **Journal of Personality and Social Psychology,** 82: 835–848.

Dovidio, J. F., & Gaertner, S. L. 1981. The effects of race, status, and ability on helping-behavior. *Social Psychology Quarterly,* 44: 192-203.

Dovidio, J. F., & Gaertner, S. L. 2000. Aversive racism and selection decisions: 1989 and 1999. *Psychological Science,* 11: 315-319.

Dubois, R. W., Alexander, C. M., Wade, S., Mosso, A., Markson, L., Lu, J. D., Nag, S., & Berger, M. L. 2002. Growth in use of lipid lowering therapies: Are we targeting the right patients? *American Journal of Managed Care*, 8: 81-86.

Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. 1992. Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin,* 111: 3-22.

Eagly, A. H. & Karau, S. J. 2002. Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109: 573-598.

Feinglass, J., & Salmon, J. W. 1990. Corporatization of medicine: the use of medical management information systems to increase the clinical productivity of physicians. *Journal of Health Services*, 20: 233-252.

Fernandez, J. P. 1981. *Racism and sexism in corporate life*. Lexington, MA: Heath.

Figueroa, E. B., & Woods, R. A. 2007. Industry output and employment projections to 2016. (2007). *Monthly Labor Review,* 130: 53-85.

Frazer, J. G. 1890/1959. *The golden bough: A study in magic and religion*. New York: Macmitian. (Reprint of 1922 abridged edition, edited by T. H., Gaster; original work published 1890).

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

Gerstein, H. C., Yusuf, S., Mann, J. F. E., Hoogwerf, B., Zinman, B., Held, C., Fisher, M., Wolffenbuttel, B., Bosch, J., Richardson, L., Pogue, J., & Halle, J. P. 2000. Effects of ramipril on cardiovascular and microvascular outcomes in people with diabetes mellitus: Results of the HOPE study and MICRO-HOPE substudy. *Lancet*, 355: 205-212.

Giannopoulos, G., Conway, M., & Mendelson, M. 2005. The gender of status: The laypersons' perception of status is gender-typed. *Sex Roles*, 53: 795-806.

Gilovich, T., Griffin, D. W., & Kahneman, D. 2002. *Heuristics of judgment: The psychology of intuitive judgment.* Cambridge, U. K.: Cambridge University Press.

Gladwell, M. 2000. *The Tipping Point: How Little Things Can Make a Big Difference*. Little Brown Publishing.

Goldin, C., & Rouse, C. 2000. Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians. *American Economic Review*, 90: 715-741.

Greenhaus, J. H., Parasuraman, S., & Wormley, W. M. 1990. Effects of race on organizational experiences, job performance evaluations, and career outcomes. *Academy of Management Journal*, 33: 64-86.

Greenwald, A. G., & Banaji, M. R. 1995. Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review,* 102: 4-27.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. 2003. Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology,* 85: 197-216.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*. In press.

Gruca, T. S., & Rego, L. L. 2005. Customer satisfaction, cash flow, and shareholder value. *Journal of Marketing,* 69: 115-130.

Gupta, S., & Zeithaml, V., 2006. Customer metrics and their impact on financial performance. *Marketing Science,* 25: 718-739.

Haas, J. S., Cook, E. F., Puopolo, A. L., Burstin, H. R., Cleary, P. D., Brennan, T. A. 2000. Is the professional satisfaction of general internists associated with patient satisfaction? *Journal of General Internal Medicine*,15: 122-128.

Hagan, C. M., Konopaske, R., Bernardin, H. J., & Tyler, C. L. 2006. Predicting assessment center performance with 360-degree, top-down, and customer-based competency assessments *Human Resource Management*, 45: 357-390.

Hsaio, W. C., Braun, P., Becker, E. R., & Thomas, S. R. 1988a. The resource-based relative value system. *Journal of the American Medical Association*, 258: 799-802.

Hsaio, W. C., Braun, R., Dunn, D., & Becker, E. R. 1988b. Resource-based relative values: An overview. *Journal of the American Medical Association*, 260: 2347-2353.

Isles, C. G. 2002. Patients with acute coronary syndrome should start a statin while still in hospital. *Heart*, 88: 5-6.

Ittner, C. D., & Larcker, D. F. 1998. Are non-financial measures leading indicators of financial performance? An analysis of customer satisfaction. *Journal of Accounting Research,* 36: 1-35.

Joshi, A., Liao, H., & Jackson, S. E. 2006. Cross-level effects of workplace diversity on sales performance and pay. *Academy of Management Journal,* 49: 459-481.

Judge, T. A., & Ferris, G. R. 1993. Social-context of performance evaluation decisions. *Academy of Management Journal,* 36:80-105.

Kane, J. S., Bernardin, H. J., Villanova, P. & Peyrefitte, J. 1995. Stability of Rater Leniency – 3 Studies. *Academy of Management Journal,* 38: 1036-1051.

Kleiner, K. D., Akers, R., Burke, B. L., & Werner, E. J. 2002. Parent and physician attitudes regarding electronic communication in pediatric practices. *Pediatrics*, 109: 740-744.

Kruse, K. M. 2005. *White Flight: Atlanta and the Making of Modern Conservatism*. Princeton, NJ: Princeton University Press.

Laine, C., & Davidoff, F. 1996. Patient-centered medicine: A professional evolution. *Journal of the American Medical Association*, 275: 152-156.

Landy, F. J., Shankster, L. J., & Kohler, S. S. 1994. Personnel-selection and placement. *Annual Review of Psychology,* 45: 261-296.

LaRosa, J. C., He, J., & Vupputuri, S. 1999. Effect of statins on risk of coronary disease: a meta-analysis of randomized controlled trials. *Journal of the American Medical Association*, 282: 2340–2346.

Latham, G. P., & Wexley, K. N. 1977. Behavioral observation scales for performance-appraisal purposes. *Personnel Psychology,* 30: 255-268.

Mauss, M. 1902/1972. *A general theory of magic*. New York: W. W. Norton. (R., Brain, Trans.; Original work published 1902).

McConahay, J. B. 1983. Modern racism and modern discrimination – the effects of race, racial-attitudes, and context on simulated hiring decisions. *Personality and Social Psychology Bulletin,* 9: 551-558.

McConnell, A. R., & Leibold, J. M. 2001. Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology,* 37: 435-442.

Morales, A. C., & Fitzsimons, G. J. 2007. Product contagion: Changing consumer evaluations through physical contact with "disgusting" products. *Journal of Marketing Research,* 44:272-283

Moshavi, D. 2004. He said, she said: Gender bias and customer satisfaction with phone-based service encounters. *Journal of Applied Social Psychology,* 34, 162-176.

Murphy, K. R. 1991. Criterion issues in performance appraisal research: Behavioral accuracy versus classification accuracy. *Organizational Behavior and Human Decision Processes,* 50: 45-50

National Institute for Occupational Safety and Health. 2002. *The changing organization of work and the safety and health of working people (NIOSH Publication No. 2002–116)*. Washington, DC: Author.

Negro, G., Goodman, S., & Rao, H. 2008. Red Rattinghood: Brokerage, Closure, and Negative Affiliations in Hollywood Blacklisting. *Working paper*.

Nosek, B. 2005. Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General,* 134: 565-584.

O'Leary, V. E., & Ickovics, J. R. 1992. Cracking the glass ceiling: Overcoming isolation and alienation. In U. Sekaran & F. Leong (Eds.), *Woman power: Managing in times of demographic turbulence*: 7-30. Thousand Oaks, CA: Sage.

O'Reilly, C. A., Williams, K. Y., & Barsade, S. G. 1998. Group demography and innovation: Does diversity help?" In D. Gruenfeld, B. Mannix & M. Neale, (Eds.) *Research on Managing on Groups and Teams*, pp. 183-207. Stamford, CT: JAI Press, Inc.

Pellegrin, K. L., Stuart, G. W., Maree, B., Frueh, B. C., & Ballenger, J. C. 2001. A brief scale for assessing patients' satisfaction with care in outpatient psychiatric services. *Psychiatric Services*, 52: 816-819.

Pulakos, E. D., White, L. A., Oppler, S. H., & Borman, W. C. 1989. Examination of race and sex effects on performance ratings. *Journal of Applied Psychology*, 74: 770-780.

Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. 1999. A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology,* 84: 754–775.

Ridgeway, C. L. 1991. The Social Construction of Status Value: Gender and Other Nominal Characteristics. *Social Forces*, 70: 367-386.

Roch, S. G., & O'Sullivan, B. J. 2003. Frame of reference rater training issues: Recall, time and behavior observation training. *International Journal of Training and Development,* 7: 93-107.

Rotundo, M., & Sackett, P. S. 1999. Effect of rater race on conclusions regarding differential prediction in cognitive ability tests. *Journal of Applied Psychology*, 84: 815-822.

Rozin, P., Millman, T., & Nemeroff, C. 1986. Operation of the laws of sympathetic magic in disgust and other domains. *Journal of Personality and Social Psychology, 50:* 703-712.

Rynes, S. L., Bretz, R. D., & Gerhart, B. 1991. The importance of recruitment in job choice – A different way of looking. *Personnel Psychology,* 44: 487-521.

Rynes, S. L., Heneman, H. G. III, & Schwab, D. P. 1980. Individual reactions to organizational recruiting: A review. *Personnel Psychology,* 33: 529-542.

Rynes, S. L., & Miller, H. E. 1983. Recruiter and job influences on candidates for employment. *Journal of Applied Psychology,* 68: 147-154.

Salam, S., Cox, J. F., & Sims, H. P., Jr. 1997. In the eye of the beholder: How leadership relates to 360-degree performance ratings. *Group and Organization Management*, 2: 185-209.

Schneider, B., Ehrhart, M. G., Mayer, D. M., Saltz, J. L., & Niles-Jolly, K. 2005. Understanding organization-customer links in service settings. *Academy of Management Journal,* 48: 1017-1032.

Sidanius, J., & Pratto, F. 1999. *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression*. Cambridge, U. K.: Cambridge University Press.

Simonet, D. 2005. Patient satisfaction under managed care. *International Journal of Health Care Quality Assurance*, 18: 424-440

Sixma H. J., Spreeuwenberg, P. M., & van der Pasch, M. A. 1998. Patient satisfaction with the general practitioner: a two-level analysis. *Medical Care*, 36: 212-229.

Song, M. 2004. Who's at the bottom? Examining claims about racial hierarchy. *Ethnic and Racial Studies,* 27: 859-877.

Spence, M. 1973. Job market signalling. *Quarterly Journal of Economics*, 85: 355-374.

Stasser, G., & Stewart, D. 1992. Discovery of hidden profiles by decision-making groups – Solving a problem versus making a judgment. *Journal of Personality and Social Psychology,* 63: 426-434.

Stewart, M., Brown, J. B., Donner, A., McWhinney, I. R., Oates, J., Weston, W. W., & Jordan, J. 2000. The impact of patient-centered care on outcomes. *Journal of Family Practice*, 49: 796-804.

Tabachnick, B. G., & Fidell, L. S. 2003. *Using Multivariate Statistics* (Fifth ed.). Boston: Allyn and Bacon.

Taylor, H. 2002. Patient/physician online communication: Many patients want it, would pay for it, and it would influence their choice of doctors and health plans. *Harris Interactive Health Care News*, 2: 1-3.

The Economist. 2008. Black America: Nearer to Overcoming. May 8: 34-36

Westphal, J. D., & Stern, I. 2007. Flattery will get you everywhere (especially if you are a male caucasian): Ingratiation, boardroom behavior, and demographic minority status affect the likelihood of gaining additional board appointments at U.S. companies. *Academy Management Journal*, 50: 1-22.

Wherry, R. J. Sr., & Bartlett, C. J. 1982. The control of bias in ratings: A theory of rating. *Personnel Psychology,* 35: 521-551.

Wilkinson, T. J., & Fontaine, S. 2002. Patients' global ratings of student competence. Unreliable contamination or gold standard?. *Medical Education*, 36: 1117-1121.

Willerson, J. T., & Cohn, J. N. (Eds.). 2000. *Cardiovascular medicine*. Philadelphia: Churchill Livingstone.

Wilson, G., Sakura-Lemessy, I., & West, J. P. 1999. Reaching the Top: Racial Differences in Mobility Paths to Upper-Tier Occupations. *Work and Occupations*, 26: 165-186.

Woehr, D., & Roch, S. G. 1996. Context effects in performance evaluation: The impact of ratee gender and performance level. *Organizational Behavior and Human Decision Processes*, 66: 31-41.

Yarkin, K. L., Town, J. P., & Wallston, B. S. 1982. Blacks and women must try harder: Stimulus person's race and sex attributions of causality. ***Personality and Social Psychology Bulletin***, 8: 21-24.

Yusuf, S., Sleight, P., Pogue, J., Bosch, J., Davies, R., & Dagenais, G. 2000. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. ***New England Journal of Medicine***, 342: 145-153.

TABLE 1.

Medcorp Sample Means, Standard Deviations, and Correlations between Predictor, Control and Dependent Variables[5]

		M	*s.d.*	1	2	3	4	5	6	7	8	9	10	11	12
1.	Patient Satisfaction	0.51	0.11	-											
2.	Practice Busyness	0.66	0.47	-.30	-										
3.	Full time equivalent	0.80	0.20	-.07	.11	-									
4.	Number of patients in panel	1749.77	550.63	-.10	.26	.59	-								
5.	Panel age	45.84	4.89	.07	-.07	.05	-.03	-							
6.	Chronic sickness of panel	1.04	0.12	.13	-.12	-.15	-.13	.55	-						
7.	Tenure with Medcorp (years)	14.81	8.51	.20	-.14	.14	.08	.33	-.20	-					
8.	Age (years)	50.34	6.58	.09	-.09	.16	.21	.29	-.05	.69	-				
9.	Nonwhite	.12	0.32	-.15	.02	.01	-.04	-.14	-.05	-.12	-.03	-			
10.	Female	.38	0.49	-.06	-.06	-.63	-.44	-.21	.04	-.25	-.31	.12	-		
11.	Productivity	23.00	1.97	.05	.12	.22	.30	-.06	.22	-.25	-.01	-.04	-.15	-	
12.	Quality	-0.01	1.55	.11	.03	.07	.08	.21	.05	.14	.11	.02	.04	-.03	-
13.	Accessibility to Patients	0.16	0.15	.13	-.11	-.18	-.23	.05	-.06	-.04	-.16	-.11	.21	.05	.23

---

[5] N = 113; All correlations larger than .15 are significant at p<.05

TABLE 2.
Medcorp Sample Analysis Examining Moderating Effects of Physician Race, Gender and Objective Performance on Patient Satisfaction with Physician[6]

	Model 1	Model 2	Model 3	Model 4
Controls				
Practice Busyness	-.23**	-.18*	.26**	-.22*
Full time equivalent	-.17	-.23	-.20	-.24*
Number of patients in panel	-.04	-.03	.01	.01
Panel age	-.23*	-.27*	-.20	-.24*
Chronic sickness of panel	.23*	.25*	.20	.22*
Tenure with Medcorp (years)	.34**	.35*	.38**	.39**
Age (years)	-.13	-.12	-.17	-.16
Productivity	.09	.09	.12	.14
Quality	.17*	.20*	.16	.17*
Accessibility to Patients	.15	.13	.07	.05
Female	-.22*	-.26*	-.17*	-.20*
Nonwhite	-.12	-.09	-.18*	-.17*
Interactions				
Productivity X Tenure	.06	.09	.01	.03
Quality X Tenure	.13	.13	.14	.14
Accessibility X Tenure	.12	.01	-.16	.06
Productivity X Female		.04		.00
Quality X Female		-.18**		-.18**
Accessibility X Female		-.16*		-.16*
Productivity X Nonwhite			-.18*	-.21**
Quality X Nonwhite			-.18*	-.16*
Accessibility X Nonwhite			-.13	-.10
Adjusted R^2	.17	.21	.23	.25
R^2	.25	.32	.33	.40
ΔR^2 from Model 1		.07**	.08**	.15**

TABLE 3.
Bookcorp Sample Effect of Employee Race and Gender on Customer Satisfaction with the Employee[7]

	Customer Satisfaction with the Employee					
	White Male/White Female Condition			White Male/Nonwhite Male Condition		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Nonwhite	-.25*	-.26	-.27	-.23*	-.21	-.20
Female	-.15	-.23	-.21	-.15	-.15	-.16
Age	-.09	-.05	-.05	-.22*	-.23*	-.21
Implicit Attitude Test (IAT) score	.05	.15	-.13	-.15	-.14	-.19
Woman condition (1=white woman employee, 0=white man employee)		-.28*	-.26*			
IAT Score X Woman condition			-.06			
Nonwhite condition (1=nonwhite man employee, 0=white man employee)					-.02	-.01
IAT Score X Nonwhite condition						-.28**
Adjusted R^2	.03	.08	.08	.08	.08	.14
R^2	.10	.16	.16	.14	.14	.22
ΔR^2 from previous model		.06*	.00		.00	.08**

---

[7] * $p<.05$ ** $p<.01$; N = 67 in the white man/nonwhite man condition and N = 54 in the white man/white woman condition.

TABLE 4.
Bookcorp Sample Effect of Employee Race and Gender on Customer Satisfaction with the Organizational Context[8]

	Customer Satisfaction with the Context					
	White Male/White Female Condition			White Male/Nonwhite Male Condition		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Nonwhite	-.03	-.03	-.07	-.16	-.08	-.08
Female	.09	-.03	.09	.16	.01	.01
Age	-.26	-.19	-.14	-.17	-.04	-.04
Implicit Attitudes Test (IAT) score	-.12	.03	-.04	.13	.09	.02
Woman condition (1=white woman employee, 0=white man employee)		-.45**	-.38**			
IAT Score X Woman condition			-.23*			
Nonwhite condition (1=nonwhite man employee, 0=white man employee)					-.44***	-.46***
IAT Score X Nonwhite condition						-.18*
Adjusted R^2	.00	.15	.17	.00	.18	.20
R^2	.07	.24	.29	.07	.22	.26
ΔR^2 from previous model		.17**	.04**		.15***	.04*

TABLE 5.[9]
Golfcorp Means, Standard Deviations and Correlations between Dependent, Independent and Control Variables.

		M	*s.d.*	1	2	3	4	5	6	7	8	9	10
1.	Customer Satisfaction with Facility	3.90	.22	-									
2.	Size (Number of employees)	129.05	157.51	.16	-								
3.	Average Employee Age	39.20	4.04	-.18	.09	-							
4.	Percent Temporary Employees	.04	.08	.15	-.07	.16	-						
5.	Average Customer Tenure (months)	60.66	22.10	-.15	.14	.15	.01	-					
6.	Percent Male Customers	.56	.07	.00	-.03	-.13	.06	-.03	-				
7.	Average Customer Age (years)	54.23	8.40	-.14	.17	.34	.01	.42	-.16	-			
8.	Percent Nonwhite Employees	.26	.18	.08	-.12	.11	.20	-.28	.15	.08	-		
9.	Percent Female Employees	.31	.12	-.07	-.21	.07	.05	-.07	.20	-.06	-.08	-	
10.	Quality of Capital	3.32	1.04	.08	.16	-.14	.16	-.04	.16	-.03	.10	-.08	-
11.	Facility Productivity (Profit per employee)	14614.26	7703.06	.13	.14	.12	.16	.12	.13	.16	.14	.09	.15

TABLE 6.
Golfcorp Sample Regression Results Examining the Interactive Influence of Percentage of Nonwhite Employees, Percentage of Female Employees and Objective Measures of Facility Performance on Customer Satisfaction with Facility[10]

	Customer Satisfaction with Facility			
	Model 1	Model 2	Model 3	Model 4
Controls				
Size	.21	.27*	.27*	.30**
Average Employee Age	-.21	-.24*	-.24*	-.19
Percent Temporary Employees	.18	.16	.16	.18
Average Customer Tenure	-.12	-.06	-.06	-.06
Percent Male Customers	-.06	-.08	-.08	-.11
Average Customer Age	-.09	-.13	-.13	-.10
Facility Quality	-.04	-.01	-.01	-.03
Facility Productivity	.17	.01	.01	-.05
Percent Nonwhite Employees	-.05	-.15	-.15	-.19
Percent Female Employees	.04	-.07	-.07	-.14
Interactions				
Percent Female X Quality		-.25*		-.23*
Percent Female X Productivity		-.22		-.31*
Percent Nonwhite X Quality			-.25*	-.25*
Percent Nonwhite X Productivity			-.34**	-.49**
Adjusted R^2	.00	.03	.12	.20
R^2	.13	.17	.28	.37
ΔR^2 from Model 1		.04*	.15**	.24***

---

[10] * p<.05 ** p<.01 ***p< .001; N = 66 country clubs.
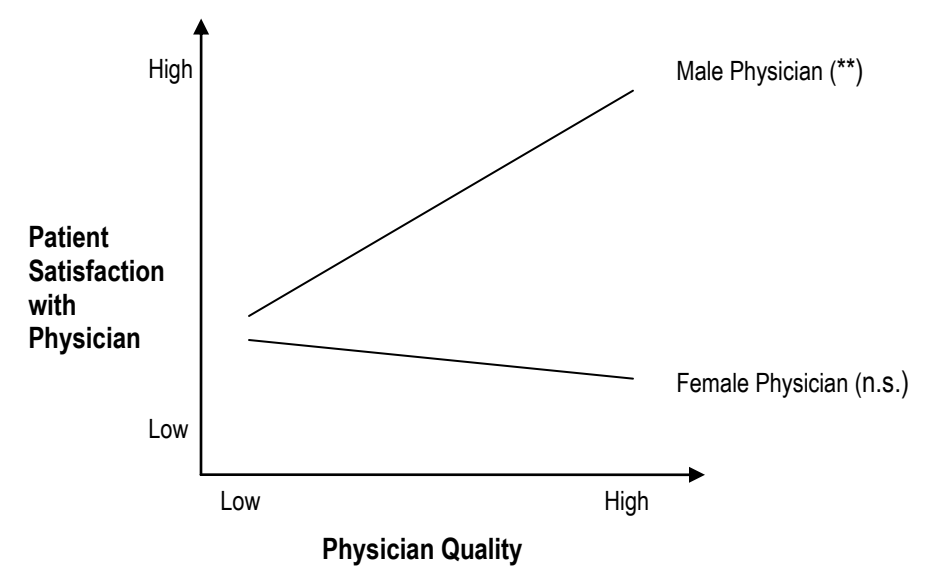
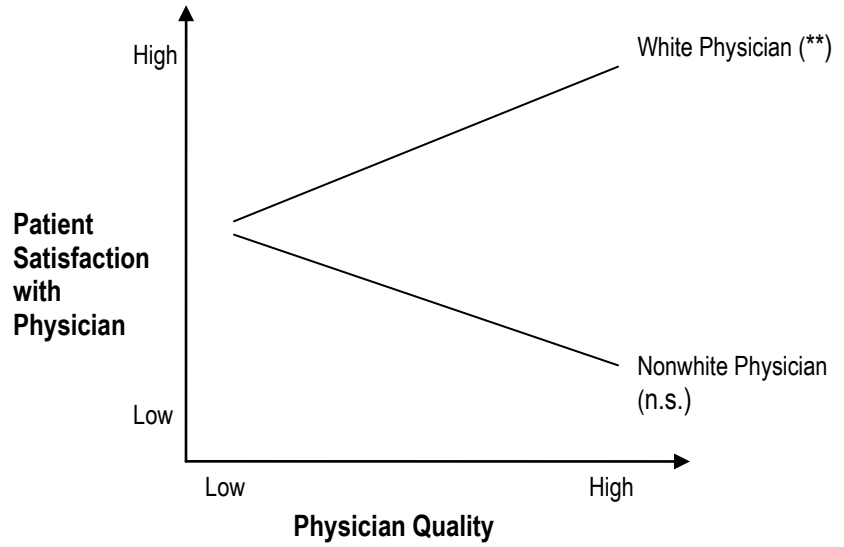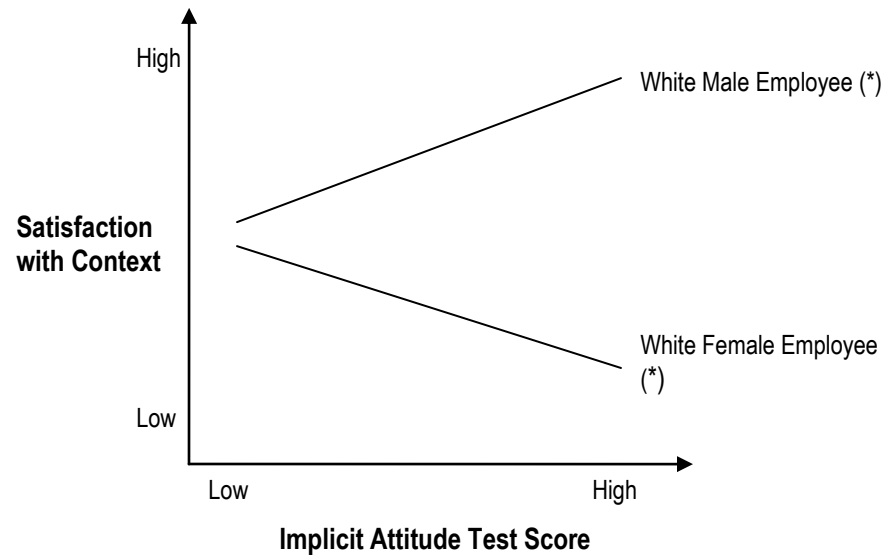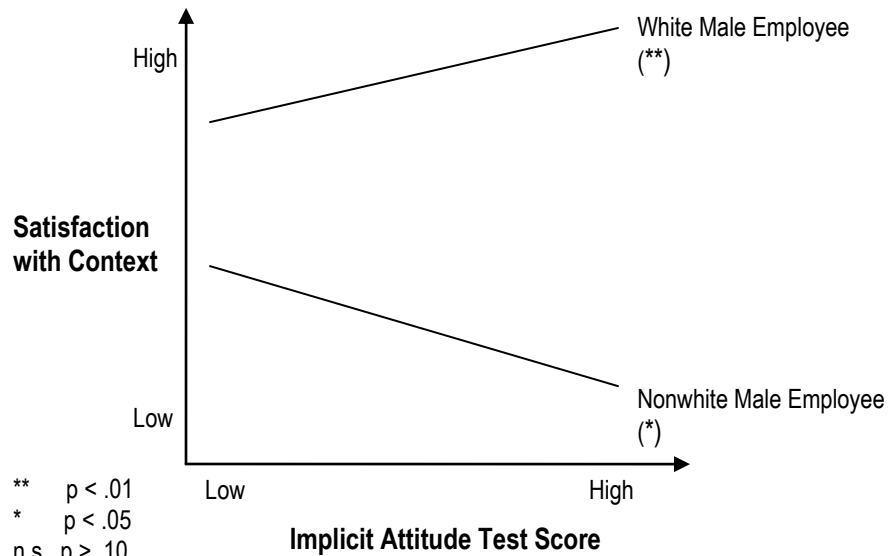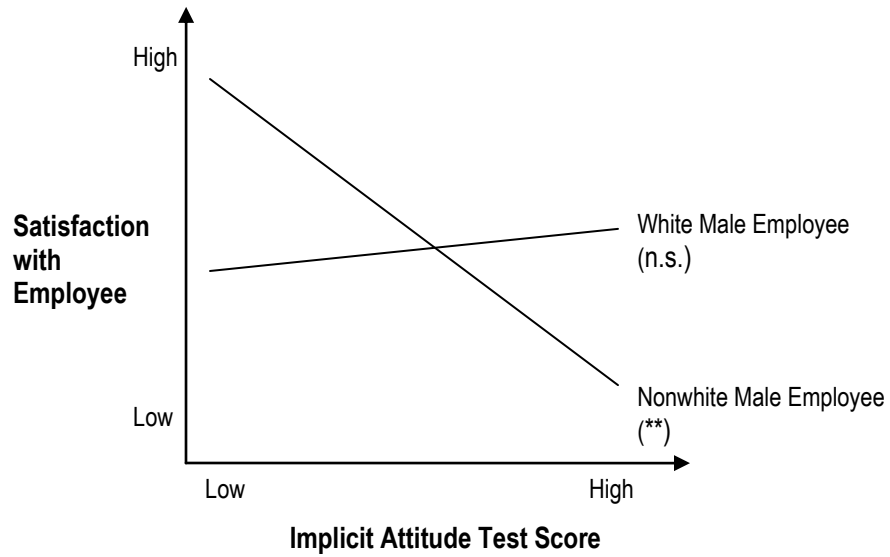**FIGURE 1.**
Conceptual Model of Customer Judgment
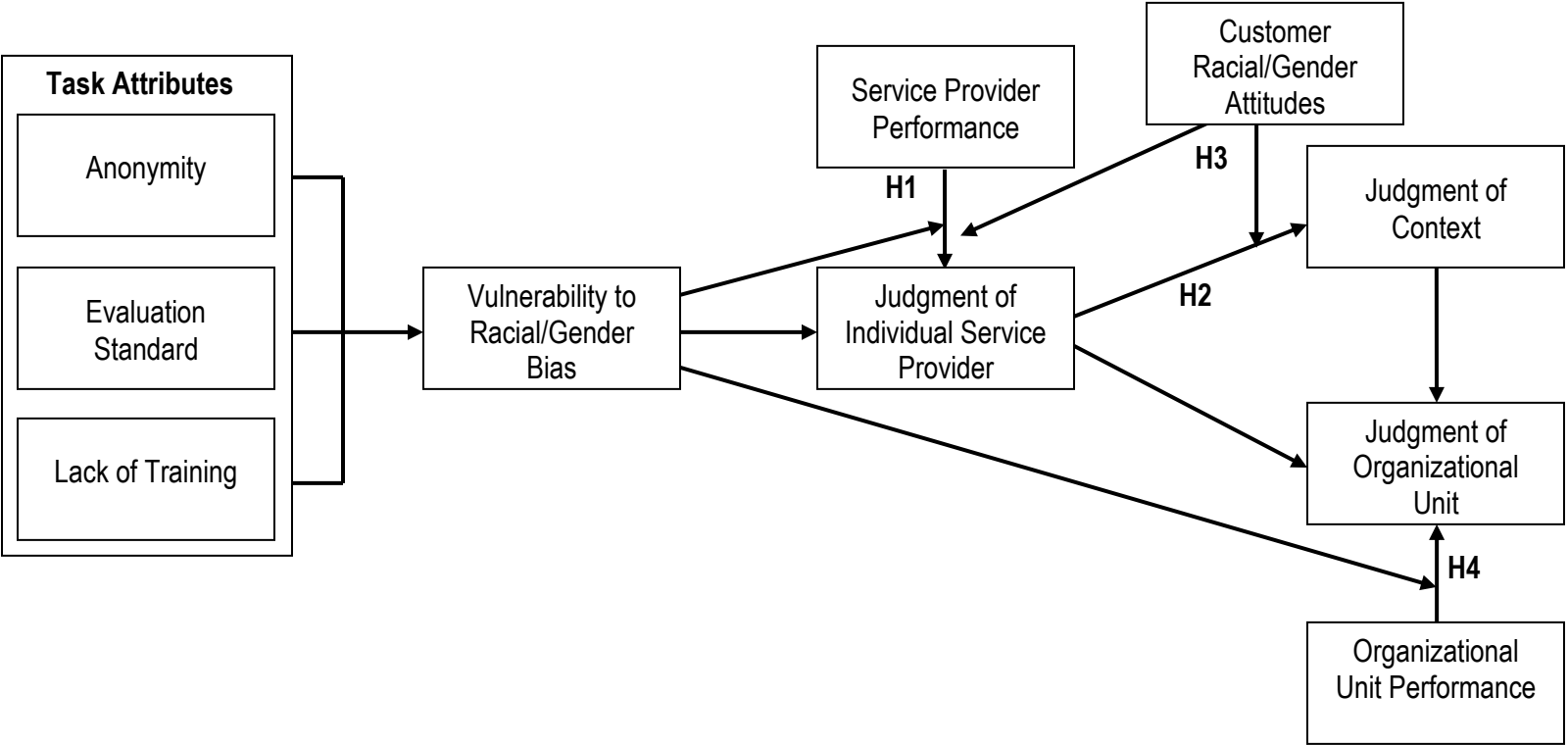
FIGURE 2.
Medcorp Sample Interactive Effects of Physician Objective Performance and Physician Demographics on Patient Satisfaction with the Physician
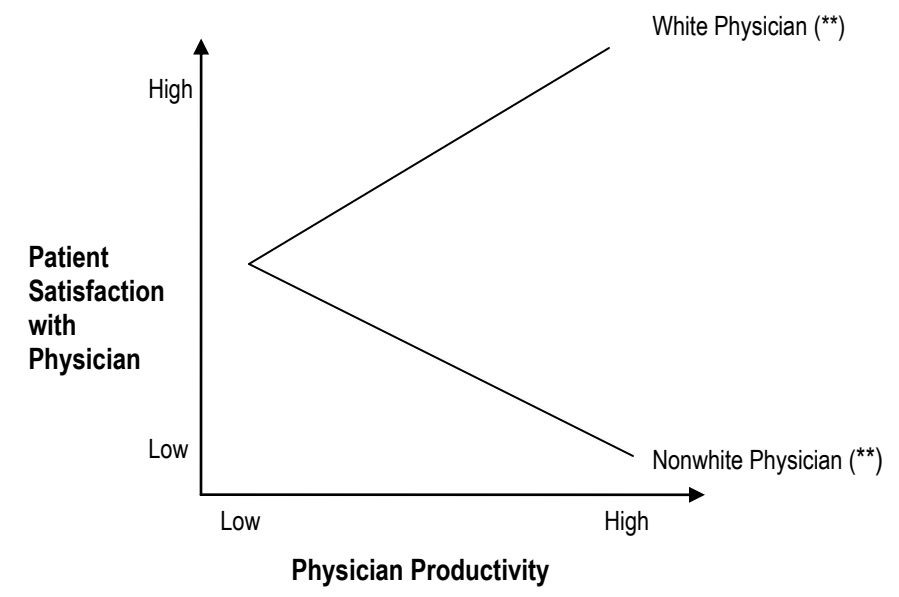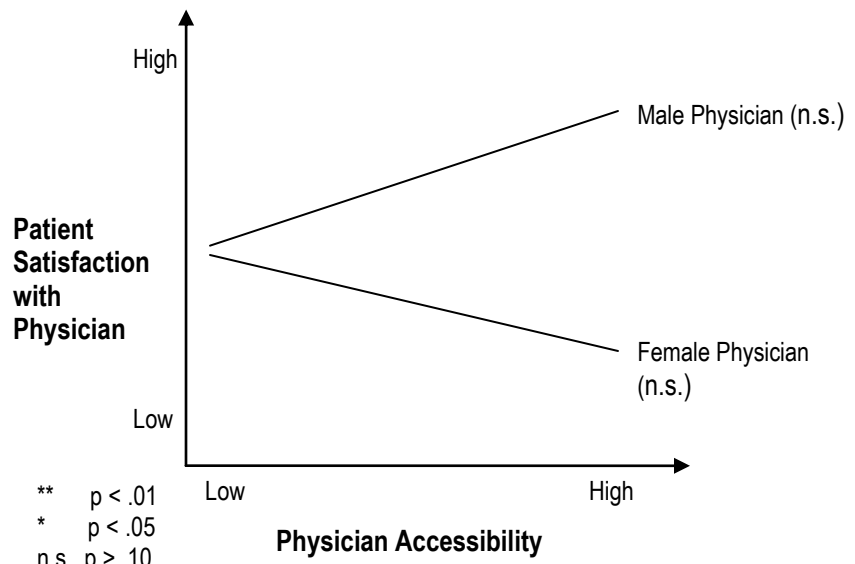
FIGURE 3.

Bookcorp Sample Interactive Effects of Employee Demographics and Customer Implicit Attitude Test Score on Customer Satisfaction with the Employee and with the Store Context

FIGURE 4.

Golfcorp Sample Interactive Effects of Percentage of White Employees, Percentage of Male Employees, Facility Quality and Facility Productivity on Customer Satisfaction with the Facility
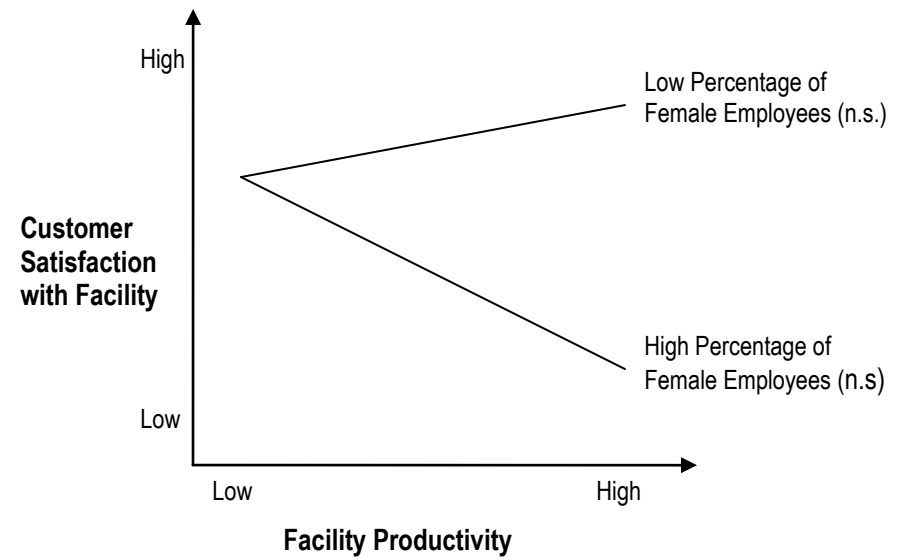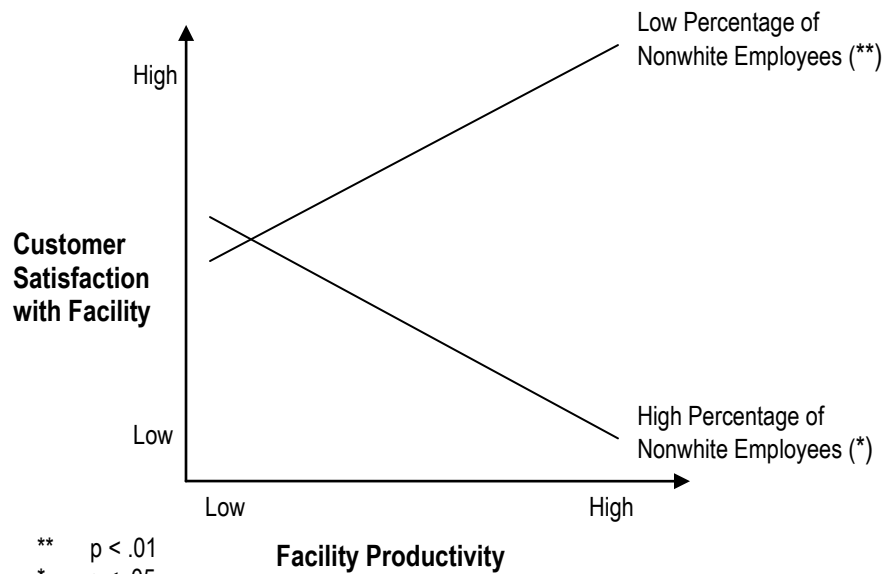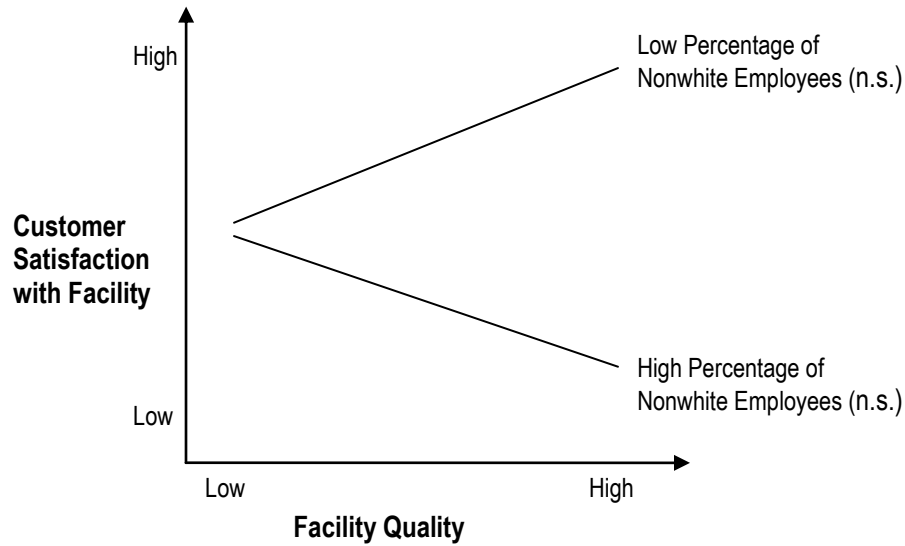
** p < .01
* p < .05
n.s. p > .10

Appendix 1: Customer Satisfaction Items Used Across the Three Studies

Medcorp Sample Customer Satisfaction with the Employee: How would you rate the following attributes of your provider (1 = very poor; 5 = excellent):
1. Attention provider paid
2. Thoroughness and competence of provider
3. Ability to ask questions of this provider

Bookcorp Sample Customer Satisfaction with the Employee: How would you rate the following (1 = very poor; 7 = excellent):
1. Speed of service
2. Quality of service
3. Availability of staff for assistance
4. Employee responsiveness to customers' issues and concerns

Bookcorp Sample Customer Satisfaction with the Context: How would you rate the following aspects of the bookstore:
1. Appearance of bookstore (1 = very poor; 7 = excellent)
2. Environment of the bookstore was conducive to learning/reading (1 = strongly disagree; 7 = strongly agree)
3. The bookstore has up to date equipment
4. This bookstore's physical facilities are visually appealing
5. The appearance of this bookstore is in keeping with the type of services provided
6. Bookstore meets expectations (1 = less than expected, 7 = better than expected)
7. Likelihood of recommending bookstore to others (1 = definitely would not, 7 = definitely would)

Golfcorp Sample Customer Satisfaction with the Facility: How would you rate the following aspects of your club (1 = very poor; 5 = very good):
1. Maintenance of grounds/Appearance of clubhouse
2. Locker rooms and Rest Rooms
3. Quality of greens
4. Condition of course
5. Pace of play
6. Condition of practice facilities
7. Ability to obtain desired tee times
8. Club meets expectations (1 = less than expected, 5 = better than expected)
9. Likelihood of recommending club to others (1 = definitely will not, 5 = definitely will)

Professor Peter A. Bamberger
William Davidson Faculty of Industrial Engineering and Management
Technion - Israel Institute of Technology
Technion City, Haifa 32000
Israel
peterb@tx.technion.ac.il

October 22, 2012

Dear Professor Bamberger,

Thank you for inviting us to revise our manuscript, "An examination of whether and how racial and gender biases influence customer satisfaction ratings," (Ms. AMJ-2008-0445) for publication consideration in *Academy of Management Journal*. Your comments and those of the anonymous reviewers were extremely helpful to us as we revised our paper. We acknowledge and very much appreciate the considerable time and energy required to produce such high quality feedback, and we have done our best to rewrite the paper in accordance with it. We believe that the paper is now much stronger as a result and hope you concur.

Our point-by-point responses to your comments and the comments of the anonymous reviewers are on the following pages. However, because the manuscript has been almost completely rewritten, we thought it might be useful to highlight some of the more significant changes below.

- Although the major purpose of our research was to demonstrate that there are racial and gender biases in customer ratings, the theory and discussion sections of our paper have been revised to focus and better support the underlying theoretical explanations of these biases. We concentrate on showing that customer satisfaction evaluations of employees, organizational contexts, and entire organizational units are biased because customers are anonymous and untrained, and because customer satisfaction judgments often lack specific evaluative standards.

- We describe supplementary analyses involving all the additional control variables you asked about. We explain that none of these other control variables changed our results but we chose to exclude some of them from the manuscript to maintain an adequate case to variable ratio. We also clarified the descriptions of each study's methods.

- We discuss several ways how the disturbing effect that we uncovered might be stopped. Specifically, we suggest bias in customer satisfaction judgments would be lessened if customers were given a specific evaluative standard and training, and if the ratings were no longer anonymous.

Again, we believe our paper has benefited enormously from your input and the input of the anonymous reviewers. If there is anything we missed or anything else you or the reviewers believe should be addressed, we very much welcome your suggestions.

1. *Bolster the theoretical contribution. More clearly develop your theory of how racial and gender biases either spread throughout or creep into organizations via customer satisfaction ratings (e.g. might accountability, anonymity, or homophily be partly responsible?). To the*

discussion add a limitations section, some rough ideas of how racial and gender biases may spread throughout organizations via customer biases, a paragraph about how your findings relate to the burgeoning literature on the diversity to performance relationship, and a paragraph on how your work contributes to relevant literature on customer service organizations and service quality.

We have done a few things here. First, we have significantly strengthened our theoretical arguments. We now more clearly describe how customer biases influence customer ratings of individual employees, of the surrounding organizational context, and the organizational unit as a whole. We have also added a figure that summarizes our conceptual arguments.

Our logic regarding how racial and gender biases creep into organizations via customer satisfaction judgments of individual employees is based on the fact that customer satisfaction judgments do not include any of the known bias-reducing safeguards common to supervisor performance ratings. Indeed customer satisfaction evaluations are anonymous and designed to elicit judgments rather than recall, which makes them highly susceptible to bias. Moreover, customers are not trained in proven bias-reduction techniques. We argue that these factors may lead customer satisfaction judgments to be highly reflective of the customer (and their biases), rather than the target.

We have also bolstered our theory regarding how bias influences customer satisfaction judgments of the organizational context. Our arguments are based on the spillover and signaling literatures. We suggest that customers tend to perceive the negative properties of low-status employees as spilling over onto the surrounding context and as signaling that organizational features, such as the context, are low-quality.

We appreciated your insight that customer accountability and anonymity and employee homophily may help explain our results. Although we mentioned customer accountability and anonymity in the previous version of the manuscript, these two concepts are now even more central to our theory. However, we do not see evidence of employee homophily as driving our results. If employee homophily were driving customer satisfaction ratings in our Golfcorp sample as you suggested, we would expect facilities with little variation in employee age or a large percentage of non-temporary employees to also receive higher customer satisfaction ratings. To test this idea, we included the interactions of the coefficient of variation of employee age (i.e. each facility's standard deviation of age/facility average age) by facility productivity and quality and percent temporary employees by facility productivity and quality in our regression models. None of these four interaction terms were significant, and they did not change our results when they were included. Adding any more variables to the Golfcorp sample regression model puts us out of compliance with the 5:1 rule (i.e. there should be at least five cases for every independent variable; Tabachnik & Fidell, 2003). Therefore, we excluded these additional five variables (i.e. coefficient of variation of employee age, four interactions) from the manuscript, but would be happy to include them if you think we should.

We have also organized our paper around theory, rather than studies. Specifically, we wove all our theoretical arguments (which were dispersed throughout in the first draft) into the front-end of the paper, and now refer to our three "studies" as "samples." We believe that these changes make the paper more palatable to an *AMJ* audience.

Finally, we have added a limitations section, ideas for minimizing the influence of customer biases on employee career outcomes, our best thoughts about how racial and gender biases may spread throughout organizations via customer biases, a paragraph about how our findings relate to the burgeoning literature on the diversity to performance relationship, and a paragraph on how our work contributes to relevant literature on customer service organizations and service quality. We agree that these additions enhance the clarity and impact of our findings.

2. THE THEORY ITSELF: *Consider including hypotheses for the main effects of employee race and gender on customer satisfaction ratings. Strengthen your arguments supporting the two-way interactions. Run three-way interactions of race X gender X objective performance in your Medcorp and Golfcorp samples.*

We now clarify that all else equal, we expect the main effects of employee race and gender to influence customer satisfaction. However, in our Medcorp and Golfcorp samples, we avoid hypothesizing these main effects and reporting our significant main effect results because they do not demonstrate the operation of bias in customer judgments. It may be that the lower customer satisfaction judgments received by members of lower status demographic groups may in fact indicate lower levels of performance relative to their higher status group counterparts or any other variables that might be strongly associated with their performance.

Perhaps it is confusing that we interpret the main effects of employee race and gender in our Bookcorp sample as evidence of customer bias, but do not interpret the main effects of employee demographics in our Medcorp and Golfcorp samples. We now clarify that the main effects of employee demographics indicate evidence of customer bias in our Bookcorp sample because the employee behavior was scripted and identical. In our field studies employees' behavior may vary, so we control for their performance before analyzing the data for bias.

As for the form of the plots, Reviewer 1 is correct that the performance appraisal literature suggests that individuals will rate high-status or low-status employee work highly when it is obviously good. We have removed this distracting sentence from our theory section because unlike supervisor performance ratings, customers rarely receive objective proof that a given employee is highly competent. In the case of physician employees (Medcorp sample), the behaviors we measured benefit patients. However, these behaviors may not universally lead to favorable patient judgments of physician competence. For example, we could imagine that some patients may interpret being highly accessible via secure e-mail as signaling that the physician has lots of free time and is not sought out by many patients and is therefore somewhat incompetent. Our argument is that customers will be more likely to make such negative judgments about employees when those employees belong to low-status demographic groups. Finally, our results and plots are very similar to another study published in *AMJ* that examines a context where judges have limited objective evidence to evaluate others (i.e. CEO board appointment recommendations; Westphal & Stern, 2007).

We tested the three-way interaction terms you suggested in our Medcorp and Golfcorp samples. However, we did not find any evidence that nonwhite women were rated lower than nonwhite men or white women. Perhaps we do not have sufficient statistical power to detect these effects.

We do not report the non-significant three-way interactions because we do not want readers to get the impression that nonwhite women do not face double jeopardy. However, we would be happy to report these non-significant findings if you feel they would add in some way to the paper.

.

3a. Clarify the methods and results sections. Suppression Effect – Relating to the Medcorp sample, 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?

A suppressor variable suppresses variance in the DV that is irrelevant to prediction of the DV (Tabachnick & Fidell, 2003). Our results are substantively unchanged regardless of whether the four variables reviewer 1 identified are included in model. Therefore, our results are not due to a suppression effect caused by those variables.

3b. 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 your Medcorp sample, 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 your Medcorp sample, Reviewer 2 (point 3c) would like to see you control for customer gender and race. Finally, in your Medcorp and Golfcorp samples, I can see a possible confound between demography/demographic composition and human capital.

We now report the interactions of objective performance and tenure. You'll see that the results are unchanged. We do not report customer gender and race in the Medcorp sample because these variables correlate too highly with physician gender and race ($r = .91$ for gender, and $r = .94$ for race). Although our results slightly improve when including these variables in the model, we do not report them due to unacceptably high multicollinearity between predictor variables (VIF = 9.1 and Tolerance = .12 for patient percent female; VIF = 10.6 and Tolerance = .08 for patient panel percent nonwhite).

In the Golfcorp sample, the organization does not collect customer race information. The organization collects customer gender information, so that is what we report in the manuscript. The Golfcorp results are unchanged regardless of whether the customer gender percentage variable is included in the model. Moreover, we found in our laboratory sample that nonwhite and women raters were no less biased in their ratings than whites or men.

3bi. In your Medcorp sample, 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.

We now more clearly note that none of the physicians in our sample were specialists. All were general practice physicians and had identical training experiences (i.e. four years of medical school followed by a three-year residency). Therefore, customer satisfaction judgments do not reflect physician specialization.

3bii. In your Golfcorp sample, 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?

We added and then struck nine control variables in response to this point: (1) average employee tenure, (2) annual employee voluntary turnover rate, and (3) annual employee involuntary turnover rate (4) whether the facility has a pool, (5) whether the facility has a fitness center, (6) whether the facility has tennis courts, (7) average employee satisfaction with their pay, (8) average employee rating of the quality of service customers receive from their facility, and (9) percent of customers who purchased the expensive "society" membership which gives customers privileges at other clubs and increases the quantity and quality of services customers receive from their home club each month. We believe these variables capture the degree to which facility leaders are trying to pursue a low-cost strategy versus a premium pricing strategy. Specifically, we believe clubs pursuing a low-cost and low-quality strategy might have higher voluntary and involuntary turnover (because of lower quality employees), lower average employee tenure, fewer amenities (i.e. tennis, fitness, pool), fewer employees who are satisfied with their pay, fewer employees thinking the facility provides a high quality of service, and fewer customers buying the expensive society membership. We ran the interactions of each of these variables by objective performance and our four race and gender interactions remained significant, but none of the nine additional main effect variables or 18 additional interactions were significant. To maintain compliance with the 5:1 rule, we do not report any of the main effects or interactions (but would be happy to include any main effects or interactions you would like to see included). We were not given access to payroll information, but Golfcorp leaders told us that average payroll costs are similar across facilities and pay grades for each position are set by leaders in the Golfcorp central offices.

3c. Correct inconsistencies and missing information in the text and tables.

We have removed the inconsistencies raised by the reviewers, added the requested missing information and removed the confusing language that Reviewer 3 noted.

3d. First run a control model including the main effects of the two demographic parameters, then separate models for the interactions with each demographic parameter, and finally a full model testing all interactions (reporting adjusted R-squared).

We now report the results the way you suggested.

3e. Are there demand effects in the Bookcorp sample? Additionally, how much satisfaction is being influenced by the videos and how much it is being influenced by prior experience in the bookstore?

Demand effects should not be an issue because our gender IAT was placed after our dependent variable (only the race IAT was placed before the dependent variable) and the results for our race and gender conditions were nearly identical. Second, there is no evidence of order effects for the IAT and dependent variables--probably because subjects still respond in socially desirable ways on the explicit measures (Greenwald, Poehlman, Uhlmann & Banaji, 2009). Finally, if the prior research is wrong and the IAT somehow did manage to create a demand characteristic, it should diminish rater bias because participants would want to avoid appearing racist or sexist.

We now clarify that the videos were taken in a bookstore at a large university on the East Coast of the U.S. and the participants were students at a large university on the West Coast of the U.S. It is very unlikely that our participants had ever seen the bookstore before.

3f. Was role-congruence a threat to external validity? You may simply need to acknowledge role-congruence as a limitation.

We wanted to be sure of our theory's external validity, too, which is why we tested it in two different field samples. There is not much evidence to suggest that a white male is congruent with the doctor role. Indeed, patients prefer their doctor to look like them (women prefer women doctors, and nonwhite patients prefer nonwhite doctors; Chen, Fryer, Phillips, Wilson & Pathman, 2005; Cooper-Patrick, Gallo, & Gonzales, 1999). All physicians had identical jobs, so we can't control for employee role-congruence. Moreover in our laboratory sample, we have no evidence indicating that a low-level bookstore employee would be a stereotypically white male job. We chose a bookstore setting for our study because we thought that if anything, it was a stereotypically female job (e.g. we consider bookstore employees to be like librarians, which some consider to be a stereotypically female job). Relatedly, our participants were likely accustomed to nonwhite and female bookstore employees. We called the bookstore and found out that over half of bookstore employees are women (74 percent) and 16 percent are nonwhite. In our Golfcorp sample, the employees had hundreds of different jobs and we do not know which ones are more congruent for white males. In sum, we list role-congruence as a possible limitation and suggest that we might expect weaker effects in work roles that are congruent with being a nonwhite or female (e.g. a bookstore).

3g. Remove references to "marginal significance."

We agree and have changed the manuscript accordingly.

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.

Thanks for these very kind words. We put a tremendous amount of effort into this research and we are glad it shows.

(1) The form of the interactions that were shown in the Medcorp and Golfcorp samples 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.

See our response to the Editor's Point 2. We have removed the distracting sentence from our theory section because unlike supervisor performance ratings, customers rarely receive objective proof that a given employee is highly competent. In the case of physician employees (Medcorp sample), the behaviors we measured benefit patients. However, these behaviors may not universally lead to favorable patient judgments of physician competence. For example, we could imagine that some patients may interpret being highly accessible via secure e-mail as signaling that the physician has lots of free time and is not sought out by many patients and is therefore somewhat incompetent. Our argument is that customers will be more likely to make such negative judgments about employees when those employees belong to low-status demographic groups. Finally, our results and plots are very similar to another study published in *AMJ* that examines a context where judges have limited objective evidence to evaluate others (i.e. CEO board appointment recommendations; Westphal & Stern, 2007).

(2) Looking at the results in Tables 1 and 2 for your Medcorp sample, 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.

See our response to Editor's point 3a. The results are substantively unchanged regardless of whether the four variables you identified are included in model, which indicates that suppressor

variables are not driving our results. Moreover, we now control for the interactions of tenure by objective performance, as the Editor requested and the results are unchanged.

(3) The "introduction" section of your Medcorp sample would be better in the general introduction to the manuscript.

We have woven our theory into the front-end of the paper and therefore have followed your request.

(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.

You are right that anonymity is important. In our theory section, we still mention accountability, but put much greater emphasis on anonymity.

(5) Can you account for double jeopardy?

The Editor picked up on this important point above (see our response to the Editor's point 2).

(6) Table 1 and the methods section for your Medcorp sample 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 and why is it the appropriate metric?

Great catches! We initially coded the white/nonwhite categories as 1 and 2, and then changed them, but forgot to update the correlation tables. The productivity variable is a composite of standardized statin prescription rate and standardized ACE inhibitor prescription rate.   We now note in the revised manuscript that the components were standardized.

We now clarify that the emails were emails from patients because all the emails were sent through a secure Internet health portal designed exclusively for patient-doctor communication. To use the system, patients logged in to a secure website that provided the patients access to their personal health records, their lab results, and a host of health related information. From this portal the patients could email their physicians health related questions. Physicians had access to outlook, and web-based emails for non-patient-related emails.

According to the National Quality Measures Clearinghouse, counting the days until the third next available appointment is the healthcare industry's standard measure of access to care and

indicates how long a patient waits to be seen. The variable was measured in whole days at the close of business each day. The final variable was each doctor's quarterly average number of days until their third available appointment slot.

(7) For your Medcorp and Golfcorp samples, enter each block separately (sex interactions, race interactions), then have both in as the third step, so you can determine the incremental validity of each over the other.

Great suggestion. We now do this.

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

We now do this.

(9) On page 19, clarify that biases against African-Americans are more or negative than those associated with Asians.

We have made this change.

(10) In your Bookcorp sample did you look at white male-white female target in one regression, and white male-black male targets in another, separate regression?

You are right that we looked at white male-white female targets in one regression and white male-black male in another, separate regression. We did not create a nonwhite-female cell because we did not find evidence of double jeopardy in our field studies. We put our two regression models side-by-side in the same table to cut down on manuscript length. We have clarified our dummy codes and variable descriptions in the new version of the manuscript.

(11) It looks like some of your items target customer satisfaction with the employee in your Bookcorp sample.

You are right and we now report the customer satisfaction with the employee results from the Bookcorp sample.

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

We have included a citation noting that there are critics of the IAT. However, much attention has been devoted to the development and validation of the IAT. Briefly, it is robust to participant faking attempts (Asendorpf, Banse, & Mücke, 2002; Banse, Seise & Zerbes, 2001; Kim, 2005; Steffens, 2004); shows acceptable internal consistency (Greenwald & Farnham, 2000), test-retest reliability (Greenwald, Nosek & Banaji, 2003), and predicts behavioral outcomes across multiple domains of social behavior (Greenwald, Poehlman, Uhlmann & Banaji, 2009). Construct validity concerns have been addressed (Greenwald, Nosek, Banaji, & Klauer, 2005), and the forthcoming meta-analysis examining the predictive validity of the IAT across domains further demonstrates the value of the approach (Greenwald, Poehlman, Uhlmann & Banaji, 2009). Moreover, this

IAT meta-analysis shows that the IAT is an especially valid predictor of race and gender attitudes (i.e. the IAT is a better predictor of intergroup discrimination behavior than self-reports of race and gender attitudes).

Asendorpf, J. B., Banse, R., & Mücke, D. 2002. Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology*, 83: 380-393.

Banse, R., Seise, J., & Zerbes, N. 2001. Implicit attitudes towards homosexuality: Reliability, validity and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie*, 48: 145-160.

Greenwald, A.G. & Farnham, SD. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, 79: 1022-1038.

Greenwald, A. G., Nosek, B., Banaji, M., & Klauer, K.C. (2005). Validity of the Salience Asymmetry Interpretation of the Implicit Association Test: Comment on Rothermund and Wentura (2004). *Journal of Experimental Psychology*, 134: 420-425.

Kim, D.Y. 2005. Voluntary controllability of the Implicit Association Test (IAT). *Social Psychology Quarterly*, 66: 83-96.

Nosek, B., Greenwald, A. G., & Banaji, M. 2007. The Implicit Association Test at age 7: A methodological and conceptual review. In J. Bargh (Ed.), *Automatic Processes in Social Thinking and Behavior*: 265–292. Psychology Press.

Steffens, M. C. 2004. Is the Implicit Association Test immune to faking? *Experimental Psychology*, 51: 165-179.

(12) Is role-congruency an issue in your Bookcorp sample?

We have no evidence indicating that a low-level bookstore employee would be a stereotypically white male job. On the contrary, we chose a bookstore setting because we thought that if anything, it was a stereotypically female job (e.g. we consider bookstore employees to be like librarians, which some consider to be a stereotypically female job). Moreover we contacted the bookstore and found out that most of the employees were women (74 percent).

(13) The Golfcorp sample methods and results sections require some editing. (a) first, is the national average of 28% in reference to the % of the population, the working population, or the working population in this industry? (b) remove repetitive information from the objective facility characteristics section (c) check that Table numbers in Text and Tables match (d) check your variable labels in Tables 4 and 5 (e) clarify why younger golf employees and older doctors might be preferred (f) do you have access to a Golfcorp employee survey?

We have made these changes. Specifically, we now clarify that twenty-eight percent of the population is nonwhite. We have removed the repetitive information and cut down the

information about quality attributes. We have corrected the table numbering and variable labels in the tables. We suggest that individuals might prefer older physicians because they have greater experience, and they might prefer younger country club employees because such employees might be more eager to please customers. Golfcorp surveyed their employees annually. We had access to the responses to this survey, and have clarified the text accordingly.

(14) Remove the phrase "marginally significant."

We have struck the term "marginal significance" from the manuscript. Even though some simple slopes are not significantly different from zero, the simple slopes are significantly different from each other. The plots are cross-over interactions and so the slopes for high-status employees are significantly more positive than the slopes for low-status employees.

(15) Get rid of Figure 3.

Figure 3 is gone.

(16) The discussion seems rather short compared to the rest of the manuscript. 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.

You'll see that we have expanded the discussion section. See the Editor's point #1.

(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.

We have removed this loaded term from the manuscript.

(18) Highlight your use of objective data as a strength, but also recognize that objective data are not automatically valid representations of the constructs of interest.

We agree and have toned this language down as well.

Thank you for your detailed comments and helpful suggestions.

R2

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.

Thank you for taking the time to say what was right about this research. We appreciate these encouraging words and we hope you will find that the paper is now even stronger.

1. Your results, across all 3 studies, are quite powerful, but your discussion section is weak. Please remove your suggestion for businesses to stop using customer satisfaction ratings in personnel decisions. Instead, highlight how organizations could transform the ratings or remove bias from the ratings. 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 please add a discussion of limitations.

The sentence suggesting that organizations stop using customer evaluations in personnel decisions is now gone. We replaced it with your idea that organizations may be able to adjust customer ratings to account for bias. We now emphasize the importance of (1) less anonymity (2) specific evaluative standards, and (3) training. Additionally we suggest that organizations could only use satisfaction surveys from frequent customers to ensure that raters have sufficient contact with targets. We also suggest that organizations could ask for customer feedback during their customer experience so that customers will be most likely to be paying attention and less likely to rely on bias when recalling their customer experience. Organizations could also give raters some "tips or pointers" that are used in training to avoid bias. Organizations might also want to let customers know how the data will be used so that customers are more motivated to judge responsibly. Another thing we mention is that organizations could insert a couple of "bias sensitive" questions in customer satisfaction judgments and "throw out" the biased customers. Alternatively, organizations may be able to statistically correct for bias when calculating customer satisfaction judgments.

We also note that different survey formats for customer rating scales might be helpful for circumventing rater biases, like forced-choice, behaviorally anchored rating scales (citing specific valued behaviors), and unweighted and weighted checklists. Also increasing the specificity of the questions and the recency of the evaluation to the exposure are other ways to eliminate bias. We now encourage organizations to consider the tradeoffs between these formats and choose the one that is most likely to reduce the effects of customer rater biases.

We have added text regarding the burgeoning literature on the diversity to performance relationship, the literature on customer service organizations and quality, and study limitations.

2. Is role-congruence an issue?

We don't believe it is, but we acknowledge role-congruence as a potential limitation in the discussion. We also call for future research to test our theory in other contexts.

1a. 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.

See our response to the Editor's point 2 above.

1b. It was difficult for me to follow your description of Relative Value Units on p.14. An example that helps to illustrate this would be helpful.

We now clarify that physicians check one of three RVU boxes after seeing each patient. If the patient appointment is a quick check-back or follow-up appointment, physicians check the first box, which is worth .5 RVU's. If the patient appointment involves at least two patient issues or concerns, but less than four, then the physician checks the middle box, indicating 1.0 RVUs for that visit. If the patient appointment involves five or more patient issues, then the physician checks the third box, which indicates 1.5 RVUs.

1c. 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).

Thank you for the helpful reference. The editor picked up on this important point in his comment 3b (see our response above), but the short answer is that we didn't include these variables in our Medcorp sample due to multi-collinearity between predictors.

2a. 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. 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.

We now more fully describe the theory for this effect in our theory section and discuss this contribution in the discussion section. We are the first to show that customer attitudes about race and gender spill over onto the store context. We greatly appreciate your suggestion to highlight the novelty of this finding in the discussion.

2b. Did students attend the bookstore that was rated?

We now clarify that the bookstore in the videos was a large East Coast University and the participants attended a large West Coast University.

2c. The manipulation check questions ask about bookstore employees (plural), which I thought was confusing.

We have dropped the manipulation check because the laboratory study was scripted and the camera angles were the same, which guaranteed that the behaviors were identical.

2d. Please justify why you don't think it is a problem that you administered the race IAT before the second video.

We do not think the IAT placement is an issue for three reasons. First, if somehow the IAT created a demand characteristic, it should create less bias because the IAT would have alerted the participants that they were in danger of appearing racist or sexist. Second, we found no evidence of a demand effect because our gender IAT was placed after our dependent variable measurement and the results for our race and gender conditions were nearly identical. Finally, there is no evidence of ordering effects for the IAT and dependent variables (Greenwald, Poehlman, Uhlmann & Banaji, 2009).

2e. For the Bookcorp sample, do you find that there are no differences for your dependent variables after the first video and before IATs?

Indeed, the customer judgments immediately after the first video and before any IATs were nearly identical to the judgments after the second video and after the IATs. This finding further confirms that the IAT did not produce any demand effect.

2f. Please plot the Bookcorp interactions and conduct simple slope analyses.

We have now done this.

2g. Fix the typo on p.26.

We have now corrected this typo.

2h. First sentence of the last paragraph on p.27 is unclear.

We have deleted the sentence.

3a. What was the average response rate for each Golfcorp facility? On what basis were customers selected?

We now clarify that the average response rate per facility was 27.3 percent (an average of 234 surveyed customers per facility). The marketing company hired to do the customer survey randomly sampled each facility's customers each quarter until they got either 20 respondents or three percent of the total customer base (whichever was larger).

3b. Did you conduct a factor analysis to verify that the Golfcorp satisfaction scale was unidimensional before creating a single scale score?

We sure did. All the items loaded on one factor when we ran a varimax rotation and our factor sorting criterion was any factor with a minimum Eigenvalue greater than one.

3c. 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?

Well, we have lots of data about job descriptions, but we aren't sure which positions are role-congruent and which ones are not, so we excluded this potentially confusing information from the paper. However, we'd love to get your thoughts. The organization has 509 unique positions from CEO to dishwasher. Two-hundred fifty six of these positions involve management responsibilities. However, only two percent of employees fill these 256 managerial positions. Forty percent of these managers are women and 12 percent are nonwhite. Finally, over half of employees work in one of the following 22 positions (ranked from most to least number of employees in each position):

		Percent Female	Percent Nonwhite
1	Server	.01	.06
2	Waitstaff	.01	.87
3	Golfcart attendant	.02	.71
4	Dishwasher	.26	.63
5	Course maintenance	.17	.08
6	Dining laborer	.15	.65
7	Line cook	.83	.14
8	Receptionist	.69	.77
9	Greenskeeper	.27	.62
10	Pro-shop marshall	.37	.44
11	Cook	.10	.04
12	Golf assistant pro	.22	.58
13	Golf captain	.01	.03
14	Locker room attendant	.05	.52
15	Housekeeper	.97	.19
16	Aerobics instructor	.17	.58
17	Busperson	.03	.75
18	Membership director	.01	.59
19	House maintenance	.16	.80
20	Golf course equipment operator	.04	.13
21	Golf groundskeeper	.47	.52
22	Pro shop starter/ranger	.48	.45

3d. Controls: 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 the Medcorp sample, I wonder why you didn't control for the race of customers in Golfcorp. If you can get this data, please include it in your analyses.

We added and then struck eight control variables: (1) whether the club has a pool, (2) whether it has a fitness center, and (3) whether it has tennis facilities, as well as (4) average customer age, (5) average customer tenure at the club, (6) percent of customers purchasing the expensive "society" membership, which allows members to have privileges at other clubs and increases the services that members receive at their home club and indicates their high socioeconomic status, (7) percent of members enrolling in the club's social program (e.g. facility managers organize bi-weekly outings and activities for social program members for a small extra charge) indicating the centrality of the club to the customers, and finally (8) the average number of members per family also indicating the centrality of the club in members' lives. Adding these eight control variables slightly strengthened our results, but we chose to exclude them to maintain an adequate case to variable ratio. Unfortunately, the organization does not and will not collect customer race data. However, we found in our laboratory sample that nonwhite and women raters were no less biased in their ratings than whites or men. Likewise, we found in our doctor sample that including customer race and gender in our model slightly strengthened our results, but we do not report these variables because of multicollinearity issues.

Minor issue 1a: The sentence on page 3, end of first paragraph is quite confusing.

We have changed this sentence.

Minor issue 1b: 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.

Thank you, this book was a very helpful reference. We now discuss how this problem might be minimized—particularly through decreased anonymity, more specific evaluative standards and increased training.

Minor issue 1c: The first sentence of your last paragraph on p.37 is confusing.

Thank you for noticing this. We have struck this sentence from the manuscript.

Thanks again for your comments and suggestions.

R3

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.

Thank you for recognizing the tremendous amount of work we put into this research. We are also extremely grateful for your ideas about ways to make the manuscript better.

1. 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.

You are right that these results are troubling and we have made note of that in the abstract. We hope these findings will bring awareness and the beginnings of resolution.

2. 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.

See our response to the Editor's point #2 above.

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 make 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.

We did not mean to argue that people haven't changed. Rather, our goal was to show that customer satisfaction surveys simply ignore everything we know about rating bias (e.g. ratings are more biased when raters are anonymous, lack specific evaluation standards and are untrained). We now mention in the discussion that it's our belief that, like some other forms of rating bias, these biases would be much weaker if managers would simply remove anonymity, create an evaluation standard, and provide customers with a minimal level of bias-reduction training. Finally, we think this paper would veer off course and be way too long if we added a fourth study examining ways to make the effect go away.

4. Thus, the paper would be much more interesting and stronger had it provided ways that show how discrimination can be minimized.

See our response to the Editors point 1. We discuss how these harmful biases could be minimized in the discussion section.

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 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.

We have bolstered the introduction and theoretical contribution by detailing how customer judgments are different from supervisor or coworker ratings. We have developed a conceptual model of customer judgments, and bolstered the theory and discussion sections. Incidentally, we have put all our theory and hypotheses up front and changed the labels of our studies from "Study 1, Study 2, and Study 3" to "Medcorp sample, Bookcorp sample, and Golfcorp sample." We think *AMJ* readers will find this format more acceptable. We also now cite more *AMJ* articles.

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 participates in the studies in comparison to the field.

Individuals may not know normative leader behaviors (especially in a lab setting), but they certainly know that racism and sexism are non-normative and therefore should be especially wary about appearing racist or sexist when they are in a controlled lab setting. Of course, social desirability is one reason we expect a smaller effect in the lab, but we also expect a smaller effect in laboratory settings because 1) raters don't know the employee, and 2) the rater was only a hypothetical customer.

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.

Thanks, and we hope they are now even stronger.

8. Would it be worthwhile to control for how often the customers see their specific physician?

Because the customer satisfaction survey is anonymous the organization does not know the visit frequency of the survey respondents. However, when organizational administrators send out the customer satisfaction surveys they send an equal number to customers who visit frequently and those who visit infrequently.

9. Could you obtain data regarding supervisor ratings and compare them to customer judgments? If so, you may be able to show that longer exposure to female and nonwhite employees limits gender and racial biases.

We agree that exposure may be an important issue that may help minimize this troubling effect in organizational contexts. We mention in the discussion that organizations might consider

placing greater weight on customer ratings coming from customers who have greater exposure to employees and contexts. However, because supervisor and customer ratings have so many other differences besides exposure (i.e. anonymity, evaluation standard, training), we aren't sure that discrepancies between supervisor ratings and customer ratings would prove the importance of exposure. Moreover, we don't have the data to test this idea. In our Medcorp and Golfcorp samples, the organizations did not collect ratings from supervisors or co-workers.

10. In the Medcorp sample the n for non-whites is very small.

We acknowledge this as a limitation but should note that we collected data from two other samples that are consistent with our Medcorp findings.

11. In the note for Table 2 – the authors should specify what IAT stands for.

Done. Thank you.

12. Table 3 – please note what correlations are significant. Currently it is not marked.

Done. Thank you.

14. The discussion is well presented but could you further develop a theory of organizational level systematic oppression and also discuss the power of customers to shape the power relationships within organizations.

In response to these suggestions, we developed a conceptual model (Figure 1) that details how customer bias might creep into organizations. We also discuss in greater detail how racial and gender biases may spread throughout organizations based on customer satisfaction judgments. Although we do not feel our data allow us to put forth a comprehensive theory of how customers shape power relationships within organizations, we think this is a great avenue for future research.

15 (minor point). Point out and highlight the strengths and the contributions of the current paper, without playing down the contribution of prior work.

We are most definitely standing on the shoulders of others. We apologize that we did not convey our tremendous respect for prior research in this area. We have changed the introduction to better reflect our admiration for these works and how they served as the foundation for our work.

Thank you for your suggestions and input.

Note: Most papers referenced in our comments above are in the manuscript's bibliography. All other full references appear at their first citation in the comments document.