

**AN EXAMINATION OF WHETHER AND HOW RACIAL AND GENDER BIASES  
INFLUENCE CUSTOMER SATISFACTION RATINGS**

**ABSTRACT**

Across three studies, we examine whether and how biases may influence customer satisfaction ratings and produce discriminatory judgments for minorities and female employees. Study 1 shows that customers are less satisfied with nonwhite and women employees than their equally-well performing white and male counterparts. Study 2 shows that raters' non-conscious racial and gender biases influence satisfaction ratings of the contextual attributes of organizations that employ nonwhite and women employees. Our final study further confirms these disturbing findings and suggests that the presence of nonwhite and women employees may produce lower aggregated customer satisfaction ratings which may ultimately hurt individuals and organizations financially. (xx words)

Customer satisfaction ratings have become a common source of performance feedback for employees and organizations (Hagan et al., 2006). Mercer Consulting Group reports that in 2006 customer ratings were of primary importance for strategic decision making and over two-thirds of organizations used such ratings to determine some aspect of employee compensation (Mercer Consulting, US Policies and Practices Report, 2007). Moreover, customer satisfaction ratings are a major predictor of organizational financial performance. 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), \$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).

Many business leaders (Bracken, Church & Timmreck, 2001) and researchers (Salam, Cox & Sims, 1997) have applauded the use of customer satisfaction ratings because they believe that aggregated evaluations are highly reliable measures of employee performance quality. However, a potential disadvantage of using customer ratings, particularly for purposes of making compensation or promotion decisions, is that customers are not immune to 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 using subjective evaluations of performance, especially when such ratings come from naïve and inexperienced raters who are not held accountable for the accuracy of their ratings (Pulakos et al., 1989; Wilkinson & Fontaine, 2002; Woehr & Roch, 1996) To date, though, surprisingly little research has examined this possibility by testing whether different biases influence customer ratings of the degree to which 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 ratings associated with race and gender are biased. We extend the existing literature on biases in supervisory ratings of employee performance by focusing on *customer satisfaction ratings*, which have rarely been examined in organizational behavior research (see Moshavi, 2004 for a rare exception). Moreover, we examine customer satisfaction ratings of not only individual employees, but also ratings of the organizational context (e.g. perceived cleanliness or appearance) as well as the organizational unit as a whole. To our knowledge, no research has examined bias in customer ratings of the organizational context or the overall organizational unit. Based on the literature on modern forms of bias (Greenwald & Banaji, 1998; Crandall & Eshelman, 2003), we contend that customer satisfaction ratings that target organizational contexts or units may be equally vulnerable to racial and gender biases as customer satisfaction ratings of employees. From an organizational perspective, finding empirical evidence of racial or gender biases in customer satisfaction ratings would suggest that there is a financial incentive for organizations to differentially and negatively treat their nonwhite or female employees. Such a finding may help explain the persistent inequality between demographic groups in the workplace.

Our research has two advantages over previous studies that have examined biases in various types of ratings. First, we take into account employees' *objectively-measured* performance behaviors when examining customer ratings. This procedure is a major methodological improvement over studies that only measure rated scores (Latham & Wexley, 1994; Landy, Shankster, & Kohler, 1994). The problem with relying solely on rated scores to assess customer bias is that such scores can be interpreted as capturing both true performance scores and biases. 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 performance ratings

that are attributable to rater 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 ratings) followed by a laboratory study that can utilize more control and enable the researcher to examine more specific questions in detail (e.g. what are some potential causes and consequences of bias in customer ratings). 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, our first study tests for bias in customer ratings in a sample of professional employees. Next, in a carefully-controlled laboratory context, we test for customer bias again, but this time in a non-professional context using book store employees. We also identify and measure some of the mechanisms that might explain the effects found in the initial study. Finally, we test for customer bias in ratings of an organization unit with a sample consisting of country clubs belonging to a large hospitality company. Since the focal unit rated shifts from the individual to the organizational level in the third study, we are able to determine whether the findings from our first two studies generalize across levels of analysis. In both field studies, customer ratings were used to determine employee salaries, which meant that any biases in ratings 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 diversity related biases on customer satisfaction ratings.

### **THEORETICAL BACKGROUND AND HYPOTHESES**

Satisfying the customer is becoming increasingly important to organizations as the global economy becomes more competitive, and 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). In a worldwide survey of 681 senior executives conducted by *The Economist* during October–December 2002, 65% of the respondents reported customers as their main focus over the next three years compared to only 18% who reported shareholders as their main focus (*The Economist*, 2003). Customer satisfaction measures are one of the four components of the balanced scorecard, which is also growing in popularity (the other three components are operational, human resources, and financial measures; Kaplan & Norton, 1992). Organizations ranging from airlines (Jenkins, 1992) to universities (Davis & Davis, 1999), use customer satisfaction ratings as a way of assessing employee and unit performance.

Obtaining high customer satisfaction ratings is critical for organizations because they are leading indicators of organizational future revenues, market share, and stock price (Ittner & Larcker, 1998). Having satisfied customers improves financial performance by increasing customer loyalty, lowering marketing costs through positive word-of-mouth advertising, reducing transactions costs, and enhancing firm reputation (e.g. Anderson, Fornell, & Lehmann, 1994; Fornell, 1992; Reicheld & Sasser, 1990). In sum, 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 employee compensation directly to customer satisfaction ratings. However, because customers are not immune to information processing and judgmental biases, these biases can potentially disadvantage

some employees or organizations if they systematically receive lower customer satisfaction ratings than would be warranted based on objective indicators of performance.

Customer ratings can focus on different aspects of organizational exchange experience. Most obviously, they can be asked to rate their server (e.g. their salesperson, teller, teacher, or physician; Haas et al, 2000; Sixma et al., 1998). This rating is personal and direct. Second, they can rate the context or environment in which they were served (e.g. the merchandise available, the newness or cleanliness of the setting, the efficiency of the technology; Pellegrin et al., 2001; Simonet, 2005). This rating is less personal and direct. And of course, they can be asked to give an overall rating 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). The latter rating should encompass opinions about both the server and the context and is therefore more global and general. We will investigate all three types of ratings.

### **Bias in Customer Satisfaction Ratings of Employees**

Research regarding bias in customer satisfaction ratings has typically focused on specific employee ratings and examined the role congruence between the employee and his or her job (Iacobucci & Ostrom, 1993; Mohr & Henson, 1996; Moshavi, 2004). The basic finding in studies taking the role congruence approach is that from the customer perspective, raters are generally less satisfied with women employees or racial minorities who occupy jobs that are normally occupied by men or Caucasians (Iacobucci & Ostrom, 1993; Mohr & Henson, 1996). These effects are taken as evidence of a pro-male or pro-Caucasian bias in performance ratings and are thought to occur because raters hold stereotypic expectations about women or minorities that are inconsistent with the attributes required for success in the particular job type being evaluated (Landy & Farr, 1980; Eagly and Karau, 2002). Meta-analytic evidence supports the idea that coworker, subordinate and supervisor ratings also reflect a role incongruence bias (Eagly, Karau, Makhijani, 1995). In addition, we believe there are

good reasons for expecting customer satisfaction ratings to be even more susceptible to racial and gender biases than ratings coming from supervisors or peers.

Unlike ratings used for internal purposes of making promotion or pay decisions about individual employees, typical customer satisfaction ratings tend to ask for affective reactions (e.g., satisfaction) or general impressions of an employee rather than focusing on specific behavioral examples of either good or poor performance. As any introductory human resource management textbook explains, asking raters to provide general impressions of an employee or to indicate their affective reactions is not a very effective way of assessing employee performance compared to ratings based on behaviors (Fisher, Schoenfeldt & Shaw, 2003). But even if a customer rating were designed to focus on behavior, a second reason why they might be more susceptible to bias than ratings from supervisors or peers is that customers are not held accountable for the ratings they report. Consequently, customers are likely to feel less responsible for providing an unbiased rating. We are not claiming that customers do not take their ratings seriously. But we think it seems reasonable to suspect that because customers will not be required to justify their ratings or provide evidence for their validity or accuracy, they will be less motivated to engage in the effortful cognitive processing required to override any automatic, non-conscious biases that may influence their ratings. A third reason why customer ratings may be prone to bias is that supervisors or peers making performance evaluations of subordinate or peer coworkers are *expected* to be objective and unbiased. It is not obvious that a similar normative expectation applies to customers. For the reasons stated above, we propose that characteristics of employees that should be irrelevant for actual job performance, such as their race or sex, are likely to influence customer satisfaction ratings.

In our studies, we tested the extent to which race and sex-based biases might influence customer satisfaction ratings across different occupations and organizational settings. Although

considerable progress has been made over the last 40 years in reducing overt expressions of prejudice and discrimination in U.S. society (Bobo, 1998), 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” (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 & Mullainathan, 2004). Similarly, when evaluators of orchestral position applicants could see the applicant’s sex they were more likely to select men. When the applicant’s sex could not be observed, the number of women hired significantly increased (Goldin & Rouse, 2000). As another example, 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 our first study, we sought to discover whether biases like those reported above might also influence customer satisfaction ratings of individual employees. We explored this possibility by examining customer satisfaction ratings of physicians reported by their patients.

## **STUDY 1**

Our prediction that patient satisfaction ratings of physicians will be susceptible to sex and racial biases is based on theoretical arguments and empirical evidence suggesting that people have



preconceived notions about others' competence based on others' membership in either a high or low status demographic group (Berger, Conner & Fisek, 1974; Berger et al., 1977). In the United States, whites and men are considered by most people to belong to members of a high status social group relative to women and ethnic minorities (see Ridgeway, 1991 for a review). This perception is not surprising given that the members of these favored groups control most of the resources and occupy most of the powerful positions in U.S. society (Sidanius & Pratto, 1999). One of the benefits of belonging to a high status social group is that observers are more likely to make favorable inferences about competence, normality, and legitimacy (Aquino & Bommer, 2003; Giannopoulos, Conway & Mendelson, 2005; Sidanius & Pratto, 1999). In contrast, members of devalued groups are subject to negative stereotypes and attributions concerning their work-related competencies (Fernandez, 1981; O'Leary & Ickovics, 1992). A possible consequence of these differences in inferential judgment is that behaviors performed by a high status group member are likely to be interpreted more favorably by observers than *the same behavior* performed by someone from a low status group. Evidence for this possibility is provided by 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 ingratiation behaviors (Westphal & Stern, 2007).

There is also evidence that racial minorities and women have to achieve higher levels of performance than whites and men to be judged as having the same level of underlying ability (Biernat & Kobrynowicz, 1997; Yarkin, Town & Wallston, 1982). Arguably, these inferences occur because women and racial minorities belong to lower status groups relative to males and whites. Compounding the adverse effects of status-based expectations is the fact that people resist information that contradicts them (Biernat & Kobrynowicz, 1997; Foschi, 2000). For example, it has been found that

people only acknowledge the competence of women if there is explicit evidence that they are doing well (Shackelford, Wood & Worchel, 1996). When no explicit, objective evidence exists, people fit observable cues with their pre-conceived expectations.

We propose that the subjective nature of customer satisfaction ratings, the lack of accountability when making ratings, and the absence of normative expectations for providing accurate, unbiased ratings, will make patient satisfaction ratings of physicians highly susceptible to status-based expectations. Based on the assertion that people tend to associate demographic status with competence and underlying ability, we expect minority and female physicians who exhibit the same *performance-related behaviors* as their white and male colleagues to receive more negative customer satisfaction ratings. The following hypothesis reflects this prediction:

*Hypothesis 1: Patients will give lower satisfaction ratings to their physicians if the physician belongs to low (i.e., women, racial minorities) rather than high (i.e., male, white)-status demographic groups, even if the physicians from each group exhibit the same level of performance-related behaviors.*

## **Sample**

Our sample was drawn from all 113 primary-care physicians (i.e., family practitioners) employed by a large health maintenance organization, hereafter referred to as Prohealth (a pseudonym). Prohealth 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. Although a powerful position, physicians are more demographically balanced than other types of powerful jobs (e.g. CEO's). The 2006 Diversity Report

by the Association of American Medical Colleges reports that 24.5% of practicing physicians are women and 12.1% are non-whites.<sup>1</sup>

## Measures

Prohealth 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 many earlier studies of performance evaluation bias because in most of these studies the only measure of employee performance was a single subjective rating. In those types of studies, there is no standard by which 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 rating.** 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

---

<sup>1</sup> 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 non-whites.

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.** Prohealth identified each physician’s race, and we coded whites “0” and ethnic minorities 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 sex.** 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%.

**Customer-centered behavior.** Patients' welfare has always been the ultimate driving force in medicine; however, physicians have historically pursued patients' welfare via a predominantly physician-centered perspective (Daghio et al., 2003; Laine & Davidoff, 1996). The patient’s point of view and concerns have historically been considered irrelevant or a hindrance to understanding the core issue—the patient’s disease. This physician-centered approach to medical care involves focusing on doctors’ convenience, rather than focusing on patients’ convenience. For example, patients historically have been expected to wait patiently for doctors, and to never keep the doctor waiting (Mishler & Waxler, 1963).

However, 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 *customer-centered*, rather than physician-centered behaviors (Laine & Davidoff, 1996; Stewart et al., 2000). We used the customer-centered behaviors identified by Prohealth as our indicator of objective physician performance.

Prohealth measures customer-centered 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. Physician compensation is tied to each of these customer-centered behaviors. Physicians who exceed the 40<sup>th</sup> percentile are given a bonus, while those below the 40<sup>th</sup> percentile are not given a bonus.

All three customer-centered behaviors are presumed to benefit customers. More productive physicians are able to treat more customer problems per visit. Therefore, higher physician productivity saves customers time because customers are able to address more health issues per visit. Higher physician accessibility saves customers unnecessary visits to the doctor by allowing customers to simply email any medical questions to their physician. Higher physician quality benefits customers by preventing costly and deadly health events such as strokes and heart attacks.

*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. Prohealth 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.

This variable 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 et al., 1988a; Hsaio et al., 1988b). According to quarterly audits by administrators, Prohealth physicians accurately record RVU's in 90 percent of patient visits. Coding errors resulting from physicians coding too many or too few RVU's 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 emails that physicians sent to patients for the quarter was used to measure another type of patient-centered behavior. Physician emails benefit patients because patients 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). Prohealth patients can send unlimited emails to their physician at no cost and Prohealth physicians are expected to reply to each email within 24 hours. Patients primarily contact their physician via email to ask basic health-related questions, to request prescription refills and to schedule follow-up appointments. Prohealth administrators encourage Prohealth physicians to contact patients via email because they think that an email might save patients a trip to the doctor, 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 et al., 2002).

The Prohealth 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 whether or not to inform their patients of their email address, and (b) they could choose whether or not to personally respond to their patients' emails. We calculated the number of emails 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 Prohealth 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).

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). According to Prohealth 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. 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 et al., 2000). To promote a higher prescription rate, Prohealth administrators send emails to physicians to remind doctors to prescribe such treatment. While prescription of these medications

benefits patients by helping patients avoid death and reduce healthcare expenses, individual physicians do not generally see their utility on a daily basis. Indeed, physicians often forget to prescribe these medications (Isles, 2002).

This 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 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 Prohealth 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. This variable is automatically calculated by the Prohealth computer server at the end of each business day and represents the average time until each physician's third available appointment for the quarter.

**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.** ProHealth 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 what type of person should be a doctor, 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 ProHealth (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 Prohealth for more years may have more loyal, satisfied patients.

## **Results**

Table 1 reports the means, standard deviations, and correlation coefficients between the dependent, independent, and control variables. To test our hypothesis and account for the actual behaviors of the employees, we examined the interaction of the objective measures of customer-centered behavior (i.e. quality, productivity and responsiveness) by the employee race and gender. 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). We entered all of the control and independent variables in Model 1. In Model 2 we entered all hypothesized two-way interactions involving sex and the three dimensions of objective performance, and in Model 3 we entered all two-way interactions involving race and the objective performance measures. The full model includes all interaction effects. Table 2 presents the results of this analysis.

-----  
Insert Tables 1 and 2 here  
-----

Hypothesis 1 suggests that for a given level of performance customer satisfaction ratings would be lower for nonwhite and female employees. The two-way sex X objective performance interactions

as a set explained a significant amount of incremental variance in the dependent variable ( $R^2 = .08$ ,  $p < .01$ ) providing preliminary support for this hypothesis. Inspection of the individual regression weights showed that the physician accessibility X sex and physician quality X sex 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 1.

-----  
Insert Figure 1 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). For male physicians the coefficient of the simple slopes of quality ( $b = .31$ ,  $p < .01$ ) and accessibility behaviors ( $b = .13$ ,  $p < .10$ ) on customer satisfaction were both significant and positive. However, for female physicians the coefficient of the simple slope of quality behaviors on customer satisfaction is negligible ( $b = -.01$ , n.s.) and for accessibility behaviors is marginally negative ( $b = -.17$ ,  $p < .10$ ).

In addition, the two-way race X objective performance interactions as a set explained a significant amount of incremental variance in the dependent variable ( $R^2 = .06$ ,  $p < .05$ ) providing preliminary support for our hypothesis. 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 1. Simple slope analysis reveals that for white physicians the coefficient of the simple slope of quality ( $b = .29$ ,  $p < .01$ ) and productivity ( $b = .15$ ,  $p < .10$ ) on customer satisfaction are significant and positive. However, for non-white physicians the simple slope of quality behaviors on customer satisfaction is marginally significant and negative ( $b = -.14$ ,  $p < .10$ ) and significant and negative for productivity behaviors ( $b = -.32$ ,  $p < .01$ ). Overall, we

find support for four of the relationships predicted in Hypothesis 1.

## **Discussion**

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

We suggest that the observed pattern of relationships indicates that customers may knowingly or unknowingly reveal their biases against nonwhite and female employees through satisfaction ratings. However, we must also consider this study's shortcomings. Although Study 1 included a large percentage of women, it only included a small percentage of nonwhites. Moreover, many of the non-whites were Asians rather than African-Americans. Biases against African-Americans may be more or less negative than those associated with Asians and so a study that included African-Americans might be better able to detect the influence of such biases on customer satisfaction ratings. 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. Third, we did not measure whether customer raters had preexisting bias against women and minorities. Finally, the physician sample in Study 1 is highly susceptible to bias due to role-incongruence effects since the prototypical physician is likely to fit the

profile of a white male. It would be informative to know whether our findings would be replicated in an occupation where the prototypical employee is not necessarily a white male. We designed Study 2 to address the limitations of Study 1. Rather than physicians, we used a type of occupation that our raters were familiar with and that would probably not be associated with a particular demographic group. 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 Study 1 and we specifically assessed racial and gender prejudice as a possible moderator of the effects of employee demographics on customer satisfaction ratings.

## STUDY 2

In Study 1, we correlated objective measures of employee behavior to patient satisfaction ratings to examine the possibility that racial and gender bias might influence the latter. In Study 2, student raters were asked to evaluate a video of an employee-customer interaction in a university bookstore and to evaluate the employee's behavior and to provide ratings of the store environment. Therefore Study 2 is different from Study 1 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 Study 2'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 ratings depending on the employee's gender or race. Second, we used a different method for detecting the influence of bias on customer ratings by assessing how student participants who were asked to assume the role of customers would evaluate the *organization* (i.e. the bookstore) in which the employee-customer interaction takes place. We used ratings of the organization to assess the influence of subtle, covert bias because the laboratory method used in Study 2 is susceptible to experimenter demand

effects. There are reasons to suspect that in the lab, with an experimenter present, participants might view themselves as being under scrutiny and would therefore be motivated to appear unbiased when rating an employee, especially a woman or African American. However, they might be less conscious of trying to appearing unbiased when asked to evaluate the store environment and so a measure of customer satisfaction with the store context might be more effective at detecting racial and gender bias.

It is not unusual for customer satisfaction ratings to ask for impressions of the organizational context (e.g. the store, club or unit with which the person has experience) rather than a specific employee (Ittner and Larcker, 1998). Thus, our method we used to detect bias in Study 2 has external validity. Based on our theoretical argument that race and gender stereotypes can influence customer ratings, we expected customers who observe an employee-customer interaction involving a woman or an African-American employee to evaluate the store environment more negatively than if the employee was a man or a White. Our prediction draws from research on the notion that the negative properties of a low-status item or person can spill over on to the nearby context (Rozin, Millman & Nemeroff, 1986). Indeed, low-status employees have 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, Goodman & Rao, 2008). Thus, we argue here that the mere presence of members of lower status groups in the store environment may lead to perceptions of reduced organizational quality.

If the spillover argument is valid, then ratings of the store environment are likely to be either enhanced or diminished depending on whether the employee who is salient within this environment belongs to either a high or low status demographic group, respectively. The following hypothesis tests this prediction:

*Hypothesis 2: People will report lower customer satisfaction ratings of the store environment 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.*

One possibility that we did not test in Study 1 is whether the effects we found would be stronger for customers who were clearly prejudiced against women or minorities. Logically, it would make sense to predict such an effect and in Study 2 we were able to test for it because we measured participants' implicit negative attitudes towards women and minorities. Implicit attitudes are non-conscious or pre-conscious attitudes that drive automatic effects on behavior (Greenwald and Banaji, 1995). Importantly, implicit attitudes appear to be better predictors of behavior than their explicit counterparts when social sensitivity concerns are high (Greenwald, Uhlmann, Poehlman, and Banaji, 2008). 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, 2008, for a meta-analysis of the predictive validity of the implicit association test). Incorporating implicit attitudes into our theorizing, we test the following hypothesis:

*Hypothesis 3: The relationship between customer satisfaction ratings of the store environment and the status of the group to which the employee belongs (high—e.g. whites, males; low—e.g. African-Americans, women) will be stronger if the customer has negative implicit attitudes toward women and minorities.*

## **Sample**

Eighty-six university students from a major northwestern 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 “employees” and “customers” were hired professional actors and the scripted interaction was filmed in the real university bookstore (before it opened in the morning). We had three conditions and 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). Although women and nonwhites tend to be slightly less biased than white males (Swim, Aiken, Hall & Hunter, 1995), both women and nonwhites were included in our samples.

### **Design**

We used a between-subjects design, to reduce the chance of priming participants that they were participating in a race or gender related study. The customer, employee behavior, and the store background were equivalent across conditions due to preexisting scripts. 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 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 for each condition. We randomly assigned the ordering of the videos within each condition and found no evidence of an ordering effect.

### **Dependent Variable**

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 (see Appendix 1 for customer satisfaction items used across the three studies). This measure was adapted from an existing customer satisfaction survey we obtained from a large organization. Coefficient alpha for this measure was .76.

### **Predictor Variables**

*Condition.* We assigned participants to one of our three conditions based on the race and gender of the employee in the videos and then created two dummy variables—one for race and one for gender. The gender dummy variable only included participants in the white male employee or white female employee condition. The race dummy variable only included participants in the white male employee or nonwhite male employee condition.

It was important to establish that participants perceived the employees in each video as behaving similarly so that we could rule out the possibility that differences in their behavior rather than simply their demographic characteristics influenced ratings of the store environment. Therefore, we created a six-item manipulation check measure of employee performance. Participants were asked to rate the employee's performance (1 = very poor or strongly disagree, 7 = excellent or strongly agree) on several dimensions (1) availability of staff for assistance, (2) responding to customers' issues and concerns, (3) you do not receive prompt services from bookstore employees<sup>R</sup> (4) bookstore employees are not always willing to help customers<sup>R</sup> (5) bookstore employees are too busy to respond to customer requests promptly<sup>R</sup> (6) employees of this bookstore do not know what your needs are<sup>R</sup>. Coefficient alpha for these items was .71 and so we aggregated the items to create a composite measure of employee behavior.

---

<sup>R</sup> Indicates item is reverse-coded



*Implicit Bias.* To measure raters' racial and gender prejudices, we administered two Implicit Attitudes Tests (IATs) designed to assess negative attitudes towards non-whites and women. The race IAT was administered after the first video and before the second video, whereas the gender IAT was administered at the conclusion of the study. 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).

### **Control Variables**

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

### **Manipulation Check**

We found no differences between conditions for our manipulation check measure of employee performance ( $p > .10$ ). Moreover the interaction of condition by negative implicit bias toward women or nonwhites did not significantly predict this measure ( $p > .10$ ). Therefore we have no evidence that raters viewed the employee behaviors differently and so we can more confidently attribute differences in ratings of the organizational context to employee demographic differences.

### **Results**

We regressed the customer ratings of the organizational context on our controls, predictors, and interaction (i.e. IAT score by condition) to determine the degree to which customer ratings of the organizational context reflected race and gender bias. Table 3 presents the regression models we used to test Hypotheses 2 and 3.

-----  
Insert Table 3 About Here  
-----

Hypothesis 2 states that people would report lower customer satisfaction ratings of the store environment when an employee in that environment belongs to a low (i.e., women, African-American) rather than high (i.e., male, white) status demographic group. Model 2 of Table 3 shows there is a significant main effect of race and gender on ratings of the store environment. Indeed, model 2 of Table 3 shows a main effect of the woman condition ( $\Delta R^2 = .17$ ;  $\underline{b} = -.45$ ) and the nonwhite condition ( $\Delta R^2 = .15$ ;  $\underline{b} = -.44$ ), suggesting that raters' conscious biases influence ratings of the organizational context.

Hypothesis 3 suggests that people would be even more likely to report lower customer satisfaction ratings of the store environment when observing an employee belonging to a low-status demographic group when the customer had negative implicit attitudes toward that group. Model 3 shows that the coefficients for the interaction term IAT Score X Race is significant and in the expected direction ( $\Delta R^2 = .04$ ;  $\underline{b} = -.18$ ). Likewise, the coefficient for the interaction term IAT Score X Gender Condition is significant and in the expected direction ( $\Delta R^2 = .04$ ;  $\underline{b} = -.23$ ). These results suggest that ratings of the organizational context are vulnerable to non-conscious biases, and therefore hypothesis 3 is supported.

## **Discussion**

We found that raters taking the customer perspective rated the employee and 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. Interestingly, we generally found that participants rated the woman employee as worse than the nonwhite employee (nonwhite male  $M = 3.0$ , white female  $M = 2.6$ ,  $p < .05$ ). Possible explanations for this finding are that (1) individuals may be generally more biased toward women than nonwhites (but this is not consistent with prior research), (2) the race IAT was administered before and the gender IAT was administered

after participants rated employees, so participants in the race condition may have been primed that this was a study where race was important, or (3) expressing bias against women employees is more normative for business students than expressing bias against nonwhites. Regardless of the explanation, the main finding is that customer satisfaction ratings of the context were lower with nonwhite and women employees compared to white male employees.

Our results from Study 2 are consistent with ideas suggesting that contemporary forms of racial and gender bias may assume covert forms. Logically, one might expect that ratings of the organizational context, which are very distal from low-status employees, would be least influenced by rater biases. However, from the perspective of aversive and modern racism theories it can be argued that it is in rating the context that customers are least likely to make an effort to suppress gender or racial biases.

The results of Study 2 and the fact that most customer satisfaction surveys are focused on ratings at the unit level (Gupta and Zeithaml, 2006) encouraged us to test our final hypothesis in a real organization with real customers. Combining hypotheses from Studies 1 and 2, we expected that for a given level of objective unit performance, customers would report lower levels of satisfaction with an entire organizational unit when it has a larger percentage of low-status employees (i.e., women, nonwhites). 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.

### **STUDY 3**

Based on our theoretical arguments and the findings of our first two studies, we test the following hypothesis in Study 3:

*Hypothesis 4: Customers will report lower satisfaction with organizational units that employ higher percentages of employees belonging to low status demographic groups (i.e., women and minorities) even if the organizational units have the same level of performance.*

## **Sample**

Our sample was drawn from a large country club organization, hereafter referred to as LeisureGroup.<sup>2</sup> LeisureGroup 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 LeisureGroup 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**

LeisureGroup 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 (race and gender) and two types of objective club performance.

**Customer satisfaction with facility.** Like many organizations, LeisureGroup 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 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 Study 2. 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 Rest Rooms, (3) Quality of

---

<sup>2</sup> LeisureGroup is a pseudonym.

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.** LeisureGroup identified the percentage of white and nonwhite employees in each facility. Across the 66 clubs in the sample, 26.5% 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%.

**Employee sex.** We also obtained this variable from LeisureGroup records. Thirty-one percent of employees in our sample were women, which, LeisureGroup leaders believe to be consistent with the leisure industry average. However, the percentage of women in our sample is lower than 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 Study 1, 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. LeisureGroup measures two dimensions which were good reflections of performance attributes. The first is facility productivity, which is each facility's standardized profit margin. The second is each facility's level of attribute quality, which is the composite of how recently each club's golf course and clubhouse were built or fully renovated. Both dimensions reflect facility characteristics that benefit customers. Facilities with more productive employees create more value—both for LeisureGroup and for

customers. Indeed, LeisureGroup 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. For both metrics, facility managers are shown how they compare to other facilities. 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.* The degree to which facilities are well-run is determined by each facility's standardized profit margin and was used to measure the degree to which employees in a facility create value for customers. Facility productivity was calculated by LeisureGroup'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. Executives told us that facilities that are more productive tend to have more satisfied customers, more reliable managers, and more consistency in hitting performance targets.

*Facility Quality Attributes.* Over time, the condition of the golf course and the clubhouse deteriorates and needs to be re-built or refurbished. Crab grass and dead spots tend to spread on golf courses, making the course's appearance unsightly and making play more difficult. Likewise, clubhouses need to be refurbished (i.e. carpeting, wall-coverings, siding, and roofing needs to be replaced) every 10 to 20 years to maintain cleanliness and functionality. LeisureGroup 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 club 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*. These variables were measured by a firm hired by LeisureCorp in the year ending six months before the dependent variable was collected.

### **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 obtained the number of employees in each club and controlled for it in our analysis.

**Average employee age.** Customers may favor younger employees, so we controlled for employee average age in the analysis.

**Percent 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 customers. We controlled for the working status of each employee (0=temporary, 1=permanent) at the time the employee completed the survey.

**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.

**Percent 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.

**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). We entered all of the control and independent variables in Model 1. In Model 2 we entered the interactions involving race and the two dimensions of objective performance, and in Model 3 we entered the interactions involving sex and the objective performance measures. The full model includes all interaction effects. Table 6 presents the results of this analysis.

-----  
Insert Tables 4 and 5 here  
-----

Hypothesis 4 suggested that customers would report lower satisfaction with organizations that employ higher percentages of employees belonging to low status demographic groups (i.e., women and minorities) even if the organizations had the same level of performance. The positive relationship between facility characteristics that benefit customers and customer satisfaction with the club is



weaker for facilities with a high percentage of nonwhite employees as compared to clubs with a low percentage of nonwhite employees. The two-way race X objective performance interactions as a set explained a significant amount of incremental variance in the dependent variable ( $R^2 = .15, p < .05$ ) providing preliminary support for our hypothesis. Inspection of the individual regression weights showed that the facility attribute quality X race and facility productivity X race interactions were significant ( $p < .05$ ). We probed the pattern of the interaction by examining the simple slope of the objective performance measures for facilities with high and low percentages of nonwhite employees (Aiken & West, 1991). The results of this analysis are shown graphically in Figure 2.

-----  
Insert Figure 2 About Here  
-----

Simple slope analysis reveals that for clubs with a low percentage of nonwhite employees, the coefficient of the simple slope of facility quality on customer satisfaction is positive and marginally significant ( $b = .15, p < .10$ ) and positive and significant for facility productivity ( $b = .32, p < .01$ ). However, for clubs with a high percentage of nonwhite employees, the simple slope of quality on customer satisfaction is marginally significant and negative ( $b = -.14, p < .10$ ) and significant and negative for facility productivity ( $b = -.23, p < .05$ ). These results support hypothesis 4 for both of the race X objective performance interactions.

The two-way sex X objective performance interactions as a set explained a significant amount of incremental variance in the dependent variable ( $R^2 = .07, p < .05$ ) providing some further support for this hypothesis. Inspection of the individual regression weights showed that the facility quality X sex and facility productivity X sex interactions were significant ( $p < .05$ ). Figure 2 shows a stronger positive relationship between customer-benefiting characteristics and customer satisfaction for facilities that have a low percentage of female employees than for clubs that have a high percentage of

female employees. We calculated the significance of the simple slopes for the interactions (Aiken & West, 1991). For facilities with a low percentage of female employees the coefficient of the simple slopes of facility quality on customer satisfaction was marginally significant and positive ( $b=.13, p < .10$ ) and positive for facility productivity ( $b=.01, n.s.$ ). However, for facilities with a high percentage of female employees, the coefficients of the simple slopes of objective performance on customer satisfaction is marginally negative for both facility quality ( $b = -.14, p < .10$ ) and for facility productivity ( $b = -.11, p < .10$ ). These results support Hypothesis 4 for both of the sex X objective performance relationships.

## **Discussion**

In this third study, we found that customer-benefiting behaviors were positively related to customer satisfaction, but *only* for facilities with a low percentage of nonwhite and female employees. Similar to Study 1, we found that certain types of customer-benefiting behaviors were *negatively* related to customer satisfaction for facilities with high percentages of female and nonwhite employees. This was again an even stronger result than we anticipated.

## **OVERALL DISCUSSION**

We set out to determine if and how customer satisfaction ratings are influenced by racial and gender biases. Across three studies we found evidence that customer satisfaction ratings are susceptible to systematic and predictable racial and gender biases. Customers tended to provide lower ratings for women and nonwhite employees and for organizations that employ such employees, than for men and white employees and their employing organizations. We also found bias regardless of whether the customer satisfaction ratings targeted employees, the organizational context, or the organization as a whole. Interestingly, we found similar effects of subtle racism regardless of whether

the non-white employees were predominantly Asian (Study 1), African-American (Study 2), or Latino (Study 3).

In conducting this “full-cycle” work, we started with the most obvious question: are customer ratings of their direct server biased against women and minorities. We then took this bias idea into the laboratory to investigate it in more detail. Of specific interest was whether less direct ratings of the context in which the service takes place are also biased. 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 ratings of the organizational context and overall ratings of the unit providing the service also contained bias. Our approach is consistent with Cialdini’s (1995) argument that the best way to fully understand 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. Figure 3 depicts our conceptual model and summarizes our hypotheses.

-----  
Insert Figure 3 about here  
-----

We realize that our results may be disturbing for some, but we believe they help illuminate the persistence of demographic inequalities in organizations. For example, they may help explain the puzzle of why women and nonwhites make 25 percent less than their male and white counterparts in equivalent jobs (U.S. Census Bureau, 2006). Or why women and nonwhites are twice as likely as white men to be unemployed and underemployed (NIOSH, 2002). Or why whites and men are twice as likely to be managers or professionals. Certainly, economists have been perturbed by the persistence of demographic inequalities because they assume these inequalities will go away when managers compete for women and nonwhite applicants whose wages are 25 percent less costly than their white and male counterparts (The Economist, May 8, 2008 Issue). Our results suggest managers

will see differential levels of customer satisfaction with nonwhite and female employees, and the units in which they reside, and so from the company's perspective it makes economic sense to discriminate against nonwhite and female employees. That is, if these employees or units are seen as performing less well by customers, then these employees and units should receive fewer rewards, bonuses and promotional opportunities.

Although our results primarily offer only a diagnosis, they also offer some prescriptions for how to minimize the effects of gender and racial biases in customer ratings. Our results highlight in stark terms the necessity for managers to go against the customer ratings and their economic implications and hire and promote women and nonwhites. Evidence from famous discrimination lawsuits (e.g. Shoney's Restaurant; Abercrombie & Fitch) indicates that managers are keenly aware of the fact that some customers may prefer white and male employees. Executives in these cases admitted to deliberately discriminating against nonwhite employees in hiring and promotion decisions to enhance customer satisfaction and organizational profitability (Brief et al., 2000). Obviously, we had hoped to find that managers were wrong and customers were unbiased, but this is not what our results indicate. Our results suggest that if customer evaluations become widely and uncritically used to determine pay and promotion opportunities, the job outcomes of organizations with large percentages of women and ethnic minorities could be adversely impacted. In general, managers should try to avoid using customer satisfaction ratings in selection, compensation, and promotion decisions. When possible, managers of both individuals and organizations should use objective performance indicators to make such human resource decisions.

Our main theoretical contributions are to show that bias appears in customer satisfaction ratings, that the bias is included in ratings of the person and the context and that it can include implicit biases. These contributions are important because they help highlight the ways and reasons that biases

might appear in organizational contexts. We were somewhat disappointed that ratings of the facility and context appeared just as vulnerable to biases as ratings of employees. However, because of the social sensitivity of issues related to racial bias (Greenwald et al, 2008), expression of implicit racial bias is likely to appear in “less monitored” appraisals (such as ratings of the context and organizational unit as a whole; Chugh, 2004).

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 ratings to this standard and therefore determine whether the customer ratings of performance might be influenced by race and sex. Second, our subjective ratings 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 study, it is unsettling to find that even when customers are relatively wealthy community members, such customers may not respond favorably to organizational characteristics designed to benefit them when these organizations have a high percentage of lower-status employees. At LeisureGroup, employees at clubs with a high percentage of female and non-white employees can in fact be economically harmed by customer satisfaction evaluations because clubs who fail to achieve the target level of customer performance ratings (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 ratings explained between 15 and 24

percent of the variance in customer satisfaction ratings. 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 biases across our three samples is between medium and large.

## **Conclusions**

Over the last 40 years, women and ethnic minorities have entered occupations and attained positions of leadership in business and government that at one time were almost entirely the province of white men (Stroh, Brett & Reilly, 1992). To some observers these trends show that U.S. society has become more egalitarian, fairer, and less discriminatory since the Civil Rights Movement of the 1960s (Gross & Hook, 1978; Lynch, 1991). Yet women and ethnic minorities are still 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) and their earnings lag behind those of white men within the same occupational category (Dreher & Cox, 1996; Petersen & Morgan, 1995). Some writers have suggested that women and ethnic minorities should simply try harder (Wenneras & Wold, 1997; Yarkin et al., 1982). Yet our study findings suggest that it is an open question whether working longer hours, choosing demanding work assignments, or going above and beyond the minimal expectations of their jobs will necessarily help women and minorities achieve parity with their white male counterparts. What may be needed are other interventions that directly attempt to counteract the operation of stereotypic biases and unconscious beliefs that lead others to discount or ignore these efforts.

## References

- Aiken, L. S. & West, S. G. 1991. *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage Publications.
- Anderson, E. W., Fornell, C., & Lehmann, D. R. 1994. Customer satisfaction, market share, and profitability – findings from Sweden. *Journal of Marketing*, 58: 53-66.
- 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.
- 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.
- 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. & Mullainathan, 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.
- Chugh, D. 2004. Societal and managerial implications of implicit social cognition: Why milliseconds matter. *Social Justice Research*, 17: 203-222.
- 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.
- Crandall, C. S., & Eshelman, A. 2003. A Justification-Suppression model of the expression and experience of prejudice. *Psychological Bulletin*, 129: 414-446.
- Daghio, M., Ciardullo, A., Cadioli, T., Delvecchio, C., Menna, A., Voci, C., Guidetti, P., Magrini, N., & Liberati, A. 2003. GPs' satisfaction with the doctor-patient encounter: Findings from a community-based survey. *Family Practice*, 20: 283-288.
- Davis, W. E. & Davis, D. R. 1999. The University Presidency: Do Evaluations Make a Difference? *Journal of Personnel Evaluation in Education*, 13: 119-140.
- 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.
- Dreher, G. F. & Cox, T. H. 1996. Race, gender, and opportunity: A study of compensation attainment and the establishment of mentoring relationships. *Journal of Applied Psychology*, 81: 297-308.
- 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.
- Eagly, A. H., Karau, S. J., & Makhijani, M. G. 1995. Gender and the effectiveness of leaders – a meta-analysis. *Psychological Bulletin*, 117: 125-145.



- 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.
- Fisher, C. D., Schoenfeldt, L. F., & Shaw, J. B. 2003. *Human Resource Management*. Boston: Houghton Mifflin.
- Fornell, C. 1992. A national customer satisfaction barometer – the Swedish experience. *Journal of Marketing*, 56: 6-21.
- Foschi, M. 2000. Double standards for competence: Theory and research. *Annual Review of Sociology*, 26: 21-42.
- Gerstein, H. C., Yusuf, S., Mann, J. F. E., Hoogwerf, B., Zinman, B., Held, C., Fisher, M., Wolfenbittel, 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.
- 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., Uhlmann, E., Poehlman, T. A., & Banaji, M. 2008. Predictive Validity of the IAT. *Journal of Personality and Social Psychology*, In Press.

- Gross, B. R. & Hook, S. 1978. *Discrimination in Reverse: Is Turn-About Fair Play?* New York, NY: New York University 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.
- Iacobucci, D., & Ostrom, A. 1993. Gender differences in the impact of core and relational aspects of services on the evaluation of service encounters. *Journal of Consumer Psychology*, 2: 257-286.
- 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.
- James, E. H., Brief, A. P., Dietz, J., & Cohen, R. R. 2001. Prejudice matters: Understanding the reactions of Whites to affirmative action programs targeted to benefit Blacks. *Journal of Applied Psychology*, 86: 1120-1128.
- Jenkins, K. J. 1992. Service Quality in the Skies. *Business Quarterly*, 57: 13-18.
- Kaplan, R. S., & Norton, D. P. 1992. The balance scorecard – measures that drive performance. *Harvard Business Review*, 70: 71-79.
- 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.
- Laine, C. & Davidoff, F. 1996. Patient-centered medicine: A professional evolution. *Journal of the American Medical Association*, 275: 152-156.

- Landy, F. J., & Farr, J. L. 1980. Performance rating. *Psychological Bulletin*, 87: 72-107.
- 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.
- Lynch, F. R. 1991. *Invisible victims: White males and the crisis of affirmative action*. Westport, CN: Praeger.
- 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.
- Mishler, E. G. & Waxler, N. E. 1963. Decision process in psychiatric hospitalization: Patients referred, accepted and admitted to a psychiatric hospital. *American Sociological Review*, 28: 576-587.
- Mohr, L. A. & Henson, S. W. 1996. Impact of Employee Gender and Job Congruency on Customer Satisfaction. *Journal of Consumer Psychology*, 5: 161-187.
- 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.
- 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., and Rao, H. 2008. Red Rattinhood: Brokerage, Closure, and Negative Affiliations in Hollywood Blacklisting. *Working paper*.
- Nemeroff, C., & Rozin, P. 1994. The contagion concept in adult thinking in thte United States – transmission of germs and of interpersonal influence. *Ethos*, 22: 158-186.
- 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.
- 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.
- Petersen, T. & Morgan, L. 1995. Separate and Unequal: Occupation-Establishment Segregation and the Gender Wage Gap. *American Journal of Sociology*, 101: 329-365.
- 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.
- Reicheld, F. F., & Sasser, W. E. 1990. Zero defections – quality comes to services. *Harvard Business Review*, 68: 105-111.
- Ridgeway, C. L. 1991. The Social Construction of Status Value: Gender and Other Nominal Characteristics. *Social Forces*, 70: 367-386.
- 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., and Nemeroff, C. 1986. Operation of the laws of sympathetic magic in disgust and other domains. *Journal of Personality and Social Psychology*, 50: 703-712.
- 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.
- Shackelford, S., Wood, W., & Worchel, S. 1996. Behavioral styles and the influence of women in mixed-sex groups. *Social Psychology Quarterly*, 59: 284-293.
- 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 HJ, Spreeuwenberg PM, van der Pasch MA. 1998. Patient satisfaction with the general practitioner: a two-level analysis. *Medical Care*, 36: 212-229.
- 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.
- Stroh, L. K., Brett, J. M., & Reilly, A. H. 1992. All the right stuff. *Journal of Applied Psychology*, 77: 251-260.

- Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. 1995. Sexism and racism – old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, 68: 199-214.
- 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.
- Wenneras, C. & Wold, A. 1997. Nepotism and Sexism in Peer Review. *Nature*, 387: 341-343.
- 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*, In Press.
- 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. Study 1 Means, Standard Deviations, and Correlations between Predictor, Control and Dependent Variables<sup>3</sup>**

		<i>M</i>	<i>s.d.</i>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>
<b>1.</b>	Patient Satisfaction	0.51	0.11	-											
<b>2.</b>	Practice Busyness	0.66	0.47	-.30	-										
<b>3.</b>	Full time equivalent	0.80	0.20	-.07	.11	-									
<b>4.</b>	Number of patients in panel	1749.77	550.63	-.10	.26	.59	-								
<b>5.</b>	Panel age	45.84	4.89	.07	-.07	.05	-.03	-							
<b>6.</b>	Chronic sickness of panel	1.04	0.12	.13	-.12	-.15	-.13	.55	-						
<b>7.</b>	Tenure with Prohealth (years)	14.81	8.51	.20	-.14	.14	.08	.33	-.20	-					
<b>8.</b>	Age (years)	50.34	6.58	.09	-.09	.16	.21	.29	-.05	.69	-				
<b>9.</b>	Non-white	1.12	0.32	-.15	.02	.01	-.04	-.14	-.05	-.12	-.03	-			
<b>10.</b>	Female	1.38	0.49	-.06	-.06	-.63	-.44	-.21	.04	-.25	-.31	.12	-		
<b>11.</b>	Productivity	23.00	1.97	.05	.12	.22	.30	-.06	.22	-.25	-.01	-.04	-.15	-	
<b>12.</b>	Quality	-0.01	1.55	.11	.03	.07	.08	.21	.05	.14	.11	.02	.04	-.03	-
<b>13.</b>	Accessibility to Patients	0.16	0.15	.13	-.11	-.18	-.23	.05	-.06	-.04	-.16	-.11	.21	.05	.23

<sup>3</sup> N = 113; All correlations larger than .15 are significant at p<.05

**Table 2. Study 1 Analysis Examining Moderating Effects of Physician Race, Sex and Objective Performance on Patient Satisfaction with Physician<sup>4</sup>**

	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>
<b>Controls and Main Effects</b>			
Practice Busyness	-.23**	-.17*	-.21*
Full time equivalent	-.12	-.20	-.23*
Number of patients in panel	-.05	-.05	.01
Panel age	-.24*	-.31*	-.26*
Chronic sickness of panel	.25*	.28*	.23*
Tenure with Prohealth (years)	.38**	.39*	.41***
Age (years)	-.14	-.14	-.17
Nonwhite	-.10	-.09	-.14
Female	-.18	-.22*	-.18
Productivity	.11	.10	.14
Quality	.12	.18*	.17*
Accessibility to Patients	.09	.12	.05
<b>Interactions</b>			
Productivity X Female		.01	.00
Quality X Female		-.22**	-.21**
Accessibility X Female		-.17*	-.18*
Productivity X Nonwhite			-.21**
Quality X Nonwhite			-.16*
Accessibility X Nonwhite			-.10
<b>R<sup>2</sup></b>			
	.22**	.31**	.37*
<b>ΔR<sup>2</sup> from previous model</b>			
		.09**	.06*

<sup>4</sup> †p<.10 \* p<.05 \*\* p<.01 N = 113. All participants are medical doctors. The sample consisted of 100 whites, 10 Asian or Pacific Islanders, 2 blacks, and one Native American.

**Table 3. Study 2 Effect of Employee Race and Gender on Customer Satisfaction with the Organizational Context<sup>5</sup>**

	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
<b>Controls</b>						
Nonwhite	-.03	-.03	-.07	-.16	-.08	-.08
Female	.09	-.03	.09	.16	.01	.01
Age	-.26	-.19	-.14	-.17	-.04	-.04
<b>Main Effects and Interactions</b>						
Implicit Attitudes Test (IAT) score	-.12	.03	-.04	.13	.09	.02
Reported inattention	.13	.09	.08	.24	.18	.10
Woman condition (1=white woman employee, 0=white man employee)		-.45**	-.38**			
IAT Score X Woman condition			-.23*			
Reported Inattention X Woman condition			-.31**			
Nonwhite condition (1=nonwhite man employee, 0=white man employee)					-.44***	-.46***
IAT Score X Nonwhite condition						-.18*
Reported Inattention X Nonwhite condition						-.28**
Adjusted R-squared	.00	.15	.27	.05	.18	.26
R-squared	.09	.25	.38	.12	.27	.37
Change in R-squared		.16**	.13**		.15***	.10**

<sup>5</sup> N = 33 participants assigned to the white man condition, 34 assigned to the nonwhite man condition, and 21 assigned to the white woman condition





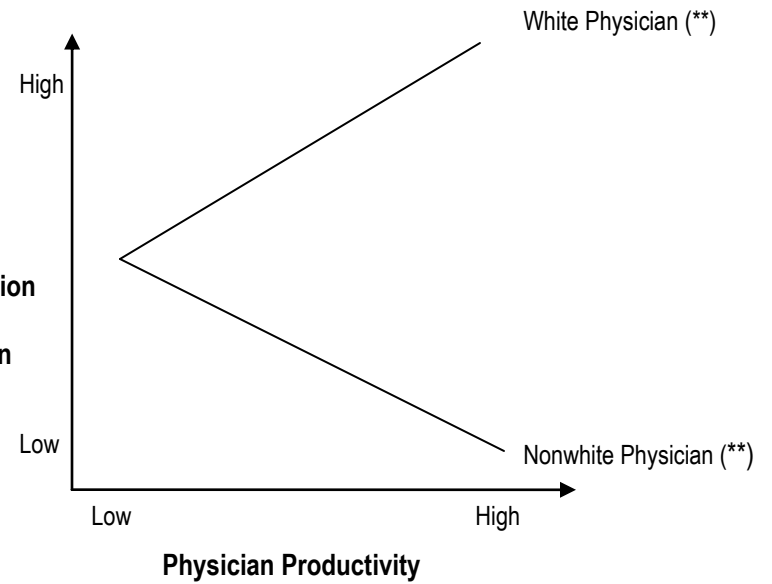
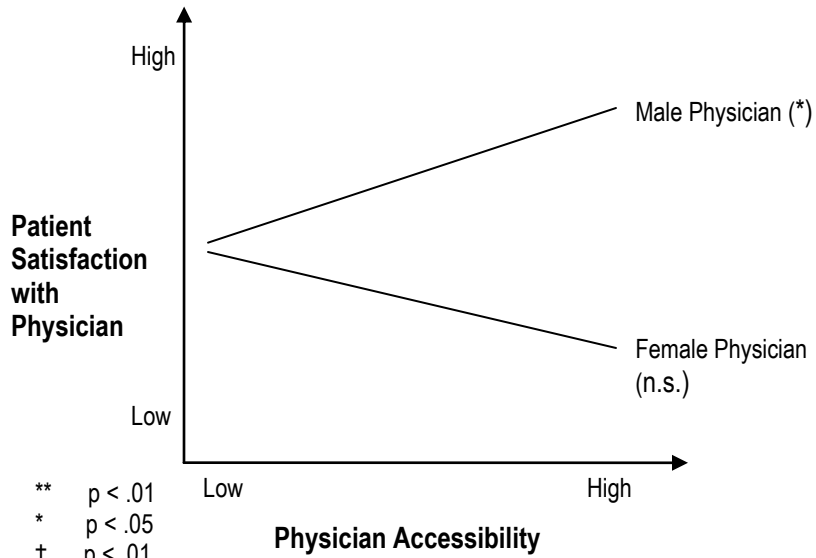
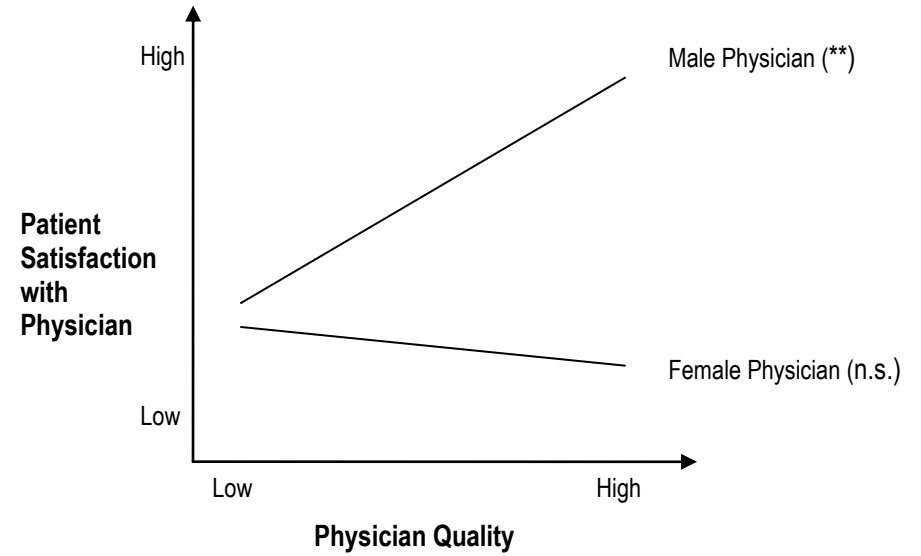
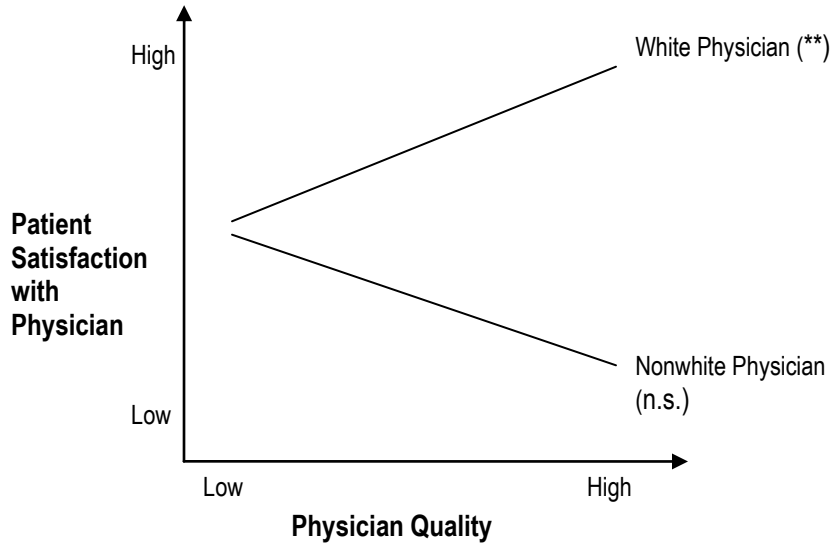
**Table 4. Study 3 Means, Standard Deviations and Correlations between Dependent, Independent and Control Variables.**

**Table 5. Study 3 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<sup>6</sup>**

	Customer Satisfaction with Facility				
	Model 1	Model 2	Model 3	Model 4	Model 5
<b>Controls</b>					
Size	.25	.21	.32*	.22	.34**
Tennis Court (1=Yes, 0=No)	.08	.16	.24	.14	.11
Fitness Center (1=Yes, 0=No)	.18	.12	.20	.14	.24
Pool (1=Yes, 0=No)	-.34*	-.42*	-.48**	-.45*	-.41*
Average Employee Age	-.15	-.12	-.13	-.08	-.10
Percent Temporary Employees	.15	.15	.13	.23	.20
Employee Voluntary Turnover Rate	.02	-.02	-.10	-.06	-.03
Employee Involuntary Turnover Rate	.17	.16	.19	.18	.18
Average Employee Pay Satisfaction	-.10	-.04	-.24	-.25	-.24
Average Employee Satisfaction with Service Provided	.36*	.35*	.32*	.31*	.33*
Average Employee Tenure	.15	.20	.22	.33*	.33*
Average Customer Tenure	-.02	-.05	-.01	-.09	-.05
Percent Male Customers	-.07	-.07	.02	-.05	.04
Average Customer Age	-.12	-.11	-.08	-.02	.00
Average Number of Customers per Family	-.07	-.04	.16	-.03	.16
Percent of Members Who Purchased Expensive Society Membership	.03	.12	.07	.16	.10
Percent of Members Who Participate in the Social Group	.03	.01	.16	.03	.22
Facility Quality	.02	.03	.02	.01	.01
Facility Productivity	.09	.08	-.04	.09	.01
<b>Main Effects</b>					
Percent Nonwhite Employees		-.08	-.24	-.07	-.31*
Percent Female Employees		-.18	-.06	-.10	-.17
<b>Interactions</b>					
Percent Nonwhite X Quality			-.28*		-.34**
Percent Nonwhite X Productivity			-.45**		-.55***
Percent Female X Quality				-.29*	-.37**
Percent Female X Productivity				-.14	-.27**
Adjusted R-Squared	.00	.05	.25	.14	.39
R-squared	.33	.36	.52	.40	.63
Change in R-squared from Model 2		<b>.03</b>	<b>.16**</b>	<b>.07*</b>	<b>.27***</b>

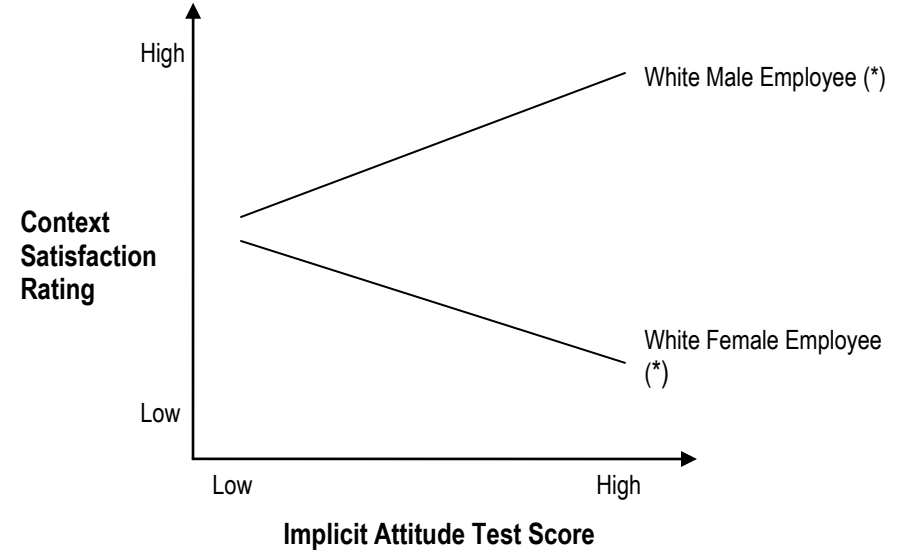
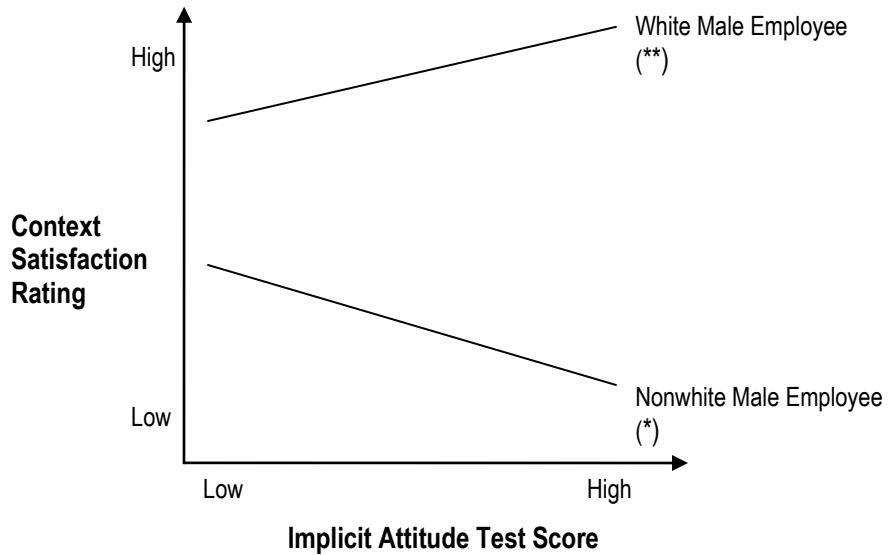
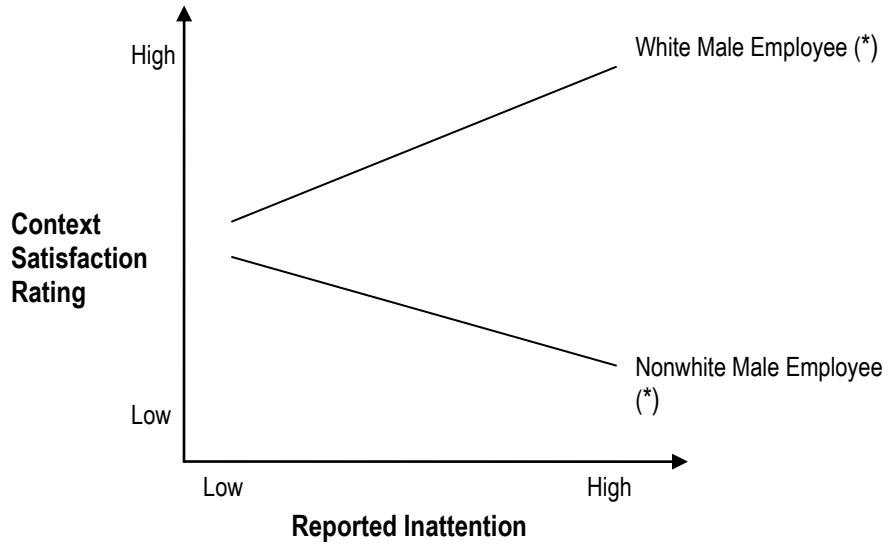
<sup>6</sup> N = 66 country clubs.

**Figure 1. Medcorp sample interactive effects of physician objective performance and physician demographics on patient satisfaction with the physician**

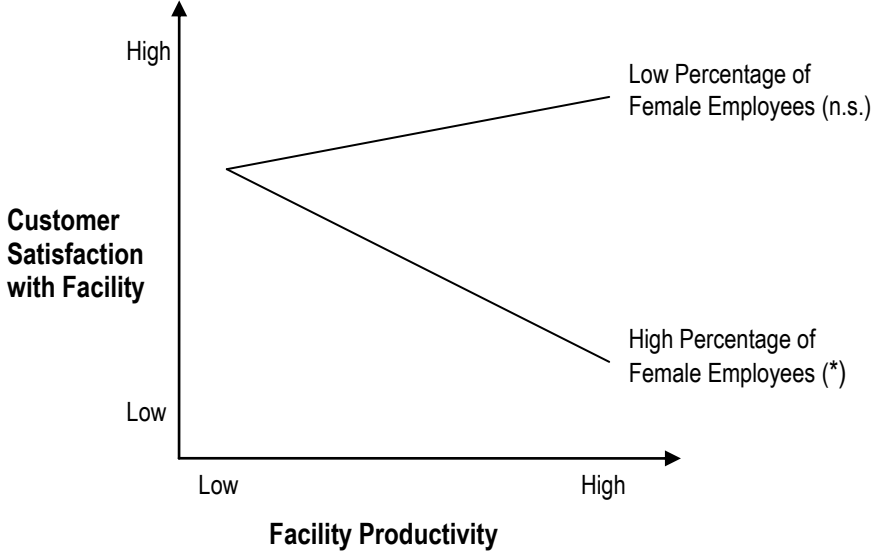
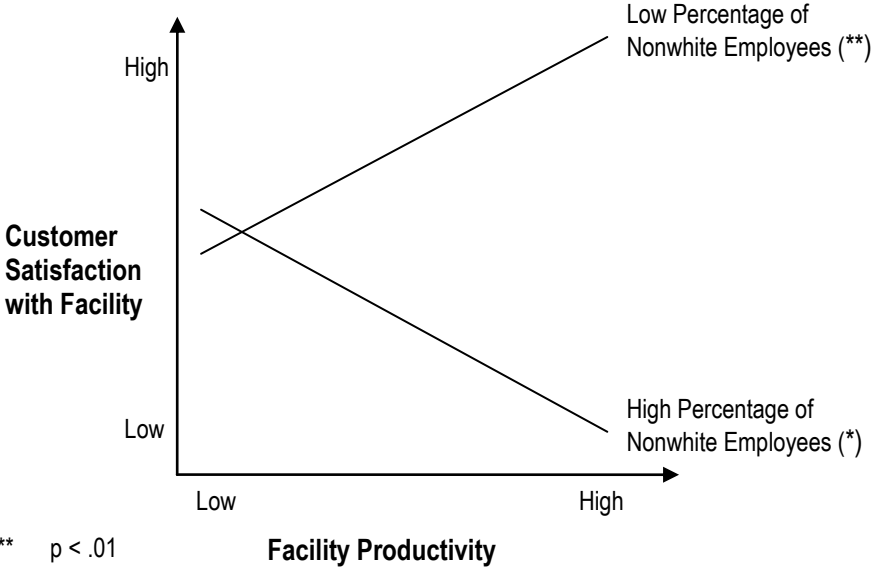
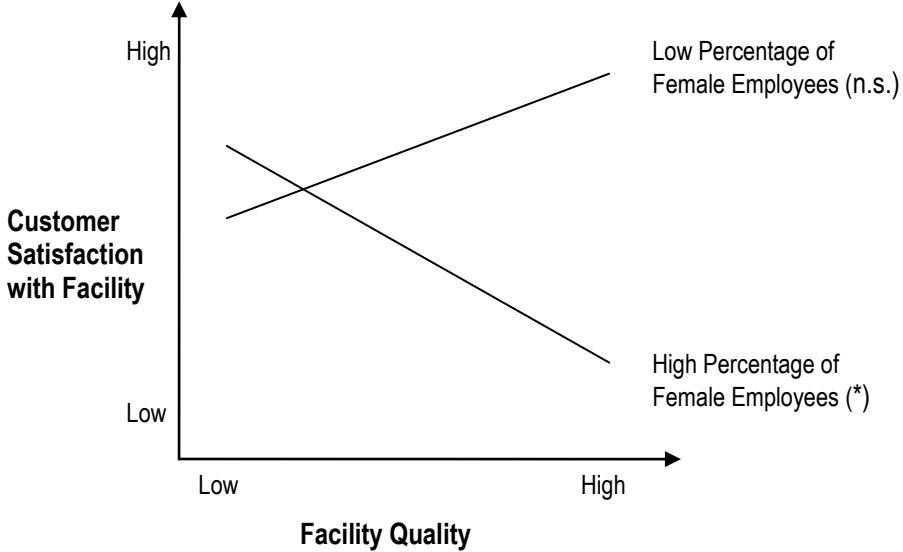
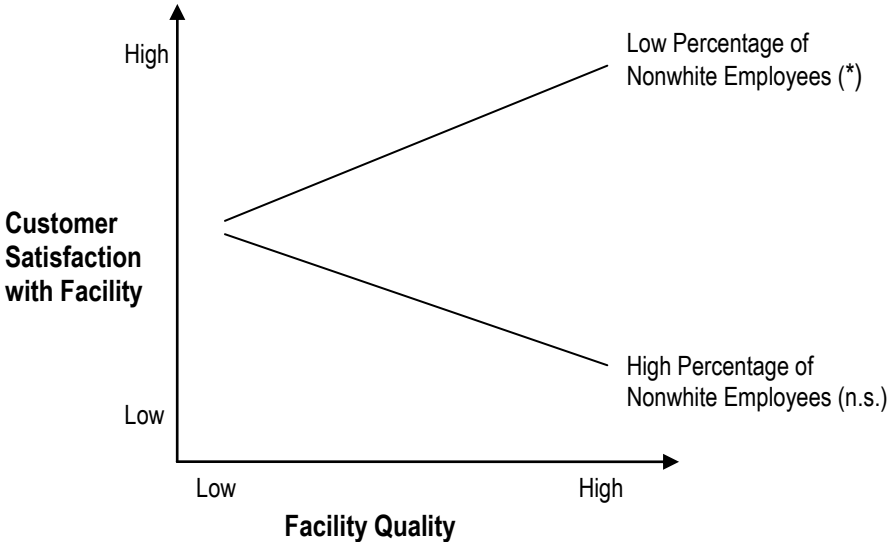


\*\* p < .01  
 \* p < .05  
 † p < .01  
 n.s. p > .10

**Figure 2. Bookcorp sample interactive effects of white male/female employee condition, white/nonwhite male employee condition, implicit attitude test score, and reported inattention on rater satisfaction with the context**



**Figure 3. Funcorp 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  
 † p < .01  
 n.s. p > .10

## **Appendix 1: Customer Satisfaction Items Used Across the Three Studies**

Study 1 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

Study 2 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)

Study 3 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)