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

DAVID R. HEKMAN
University of Wisconsin–Milwaukee

KARL AQUINO
University of British Columbia

BRADLEY P. OWENS
University of Michigan

TERENCE R. MITCHELL
University of Washington

PAULINE SCHILPZAND
KEITH LEAVITT
U.S. Military Academy at West Point

We examined whether and how various biases may influence customers' satisfaction evaluations and produce discriminatory judgments for minority and female service employees. We argue that customer satisfaction evaluations are biased because they are anonymous judgments by untrained raters that usually lack an evaluation standard. Laboratory and field samples provide disturbing evidence generally confirming our arguments and suggesting that the presence of nonwhite and women service employees may produce lower aggregated customer satisfaction evaluations that may ultimately hurt individuals and organizations financially.

Customer satisfaction surveys have become a common source of performance feedback for employees and organizations (Hagan, Konopaske, Bernardin, & Tyler, 2006). The Mercer Consulting Group (2007) reported that in 2006 customer satisfaction surveys were of primary importance for strategic decision making, and over two-thirds of organizations used such surveys to determine some aspect of employee compensation. Moreover, customer satisfaction is an important predictor of a wide range of financial measures (see Gupta and Zeithaml [2006] for a review), so it is not surprising that some companies tie some portion of employee compensation directly to customer satisfaction. For example, a 1 percent change in customer satisfaction for an average *Fortune* 500 firm has been shown to lead to a 1.02 percent change in Tobin's *q*, which equates to a change of $275 million in firm value (Anderson, Fornell, & Mazvancheryl, 2004), a

$55 million gain or loss in cash flow in the next year (Gruca & Rego, 2005), and a 5.03 percent change in return on investment (Anderson & Mittal, 2000). Satisfying customers is also increasingly important to organizations as the global economy becomes more service-oriented. Macroeconomic trends indicate 76 percent of U.S. employees work in service industries, and by 2016 the number of employees working in these industries is expected to increase by over 17 million (Figueroa & Woods, 2007). The expanding service sector is perhaps one reason why 65 percent of 681 senior executives surveyed by the *Economist* Intelligence Unit from October through December 2002 reported customers as their main focus over the next three years; only 18 percent of these executives reported shareholders as their main focus.

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

ments. As a result, they are vulnerable to biases, including those based on the "bandwagon effect," confirmation of preexisting beliefs, education or cognitive ability, and stereotypes based on the race or gender of the person being rated (Gilovich, Griffin, & Kahneman, 2002). Some researchers suspect that biases are unavoidable when gathering subjective evaluations of performance, especially when such judgments come from naïve and inexperienced raters who are not held accountable for the accuracy of their ratings (Pulakos, White, Oppler, & Borman, 1989; Wilkinson & Fontaine, 2002; Woehr & Roch, 1996). To date, though, surprisingly little research has examined how and if different biases influence the judgments of customers about whether an organization's employees, organizational attributes, or services meet or exceed their expectations.

The purpose of our research is to examine whether and how customer satisfaction ratings are potentially influenced by gender and racial bias. We extend the existing literature on biases in supervisory ratings of employee performance by focusing on *customer satisfaction* ratings, which have been mostly excluded from organizational behavior research (see Moshavi [2004] for a rare exception). We conceptualize these satisfaction ratings as judgments and examine judgments made not only about individual employees, but also about their organizational context (e.g., its perceived cleanness or appearance) and their organizational unit as a whole. To our knowledge, no prior research has examined bias in customer judgments of organizational context or overall work unit. Drawing on the literature on modern forms of racial and gender bias (Crandall & Eshelman, 2003; Greenwald & Banaji, 1995), we contend that customer satisfaction with organizational contexts or units may be vulnerable to such biases. Finding empirical evidence of racial or gender biases in customer satisfaction would suggest that organizations have a financial incentive to favor white or male employees. Such a finding may help explain the persistent inequality between demographic groups in the workplace.

Two methodological attributes of our research distinguished it from previous studies of bias in customer judgments of employee performance and allowed us to conduct a stronger test of the validity of our predictions. First, we took into account employees' *objectively measured* performance behaviors when examining customer judgments. The problem with relying solely on customer satisfaction scores to assess customer bias is that such scores can be interpreted as capturing both true

performance and biases (Landy, Shankster, & Kohler, 1994; Latham & Wexley, 1977). Rotundo and Sackett summarized 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" (1999: 816). Our study design allowed us to tease apart differences in satisfaction judgments that were attributable to customer bias arising from employee demographic characteristics from those that were due to objective employee performance (Greenhaus, Parasuraman, & Wormley, 1990; Pulakos et al., 1989; Wilkinson & Fontaine, 2002; Woehr & Roch, 1996).

Second, we heeded recent calls for researchers to engage in "full-cycle" research in which initial field-based findings are tested in the laboratory and then revalidated 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 researchers to investigate a broad initial question in a field setting (e.g., Is there bias in customer judgments?) and then conduct a laboratory study that can utilize more control and enable examination of more specific questions in detail (e.g., What is a potential cause and consequence of bias in customer judgments?). Finally, the investigators can move back to a field setting to confirm findings from the first two studies. The interplay of field and lab designs prescribed by the full-cycle approach fosters greater theoretical insight into the causality and generalizability of study findings.

Following the full-cycle research model, we first tested for bias in customer judgments regarding a sample of professional employees (doctors). Next, in a carefully controlled laboratory setting, we tested for customer bias again, but this time in a bookstore. We also identified and measured a specific mechanism that might explain the observed effect. Finally, we tested for customer bias in judgments of an organizational unit with a sample consisting of country clubs belonging to a large hospitality company. Since the focal unit rated shifted from the individual in the first sample to the organizational in the third sample, we were able to provide some initial confirmation of the generalizability of our theory. In the following section, we present the theoretical rationale for our predictions regarding the possible effects of customers' diversity-related biases on their satisfaction judgments.

THEORETICAL BACKGROUND
AND HYPOTHESES

Customers are often asked to assess an individual service provider, such as a salesperson, teller, teacher, or physician (Davis & Davis, 1999; Haas, Cook, Puopolo, Burstin, Cleary, & Brennan, 2000; Sixma, Spreeuwenberg, & van der Pasch, 1998); the quality of an environment in which they were served (e.g., the merchandise available, the newness or cleanness of the setting, or the efficiency of the technology) (Pellegrin, Stuart, Maree, Frueh, & Ballenger, 2001; Simonet, 2005); or the unit or group providing a service (e.g., a bank, school, country club, or law firm (Anderson et al., 2004; Ittner & Larcker, 1998). This last type of judgment is likely to include opinions about both the server(s) and the context in which the economic transaction occurred and is therefore a more global judgment than the other two types previously mentioned. In this study, we investigated the possibility of systematic bias in all three types of satisfaction judgments, thereby providing a strong test of the potential generality of such biases. An assumption of our study was that, like anyone else who makes a social judgment, customers are not immune to information-processing biases. One potential source of this bias is the demographic characteristics of the service providers who are being rated.

Racial and Gender Biases in Customer Satisfaction Judgments

U.S. society has made considerable progress in reducing overt expressions of prejudice since the Civil Rights Movement of the 1960s (Bobo, 1998). Yet despite these gains, there is abundant social-psychological evidence that biases against women and minorities persist in a more covert and nonconscious form. Researchers have used terms such as "modern racism" (McConahay, 1983), "aversive racism and sexism" (Dovidio & Gaertner, 1981; Gaertner & Dovidio, 1986), and "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 in a variety of social domains (see Brief, Dietz, Cohen, Pugh, and Vaslow [2000] for an overview). For example, Bertrand and Mullainathan (2004) 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 job interviews than applicants with white-sounding names (e.g., Kristin, Brad). Similarly, when evaluators of orchestra position applicants

could see an applicant's gender, they were more likely to select men. When the applicant's gender could not be observed, the number of women hired significantly increased (Goldin & Rouse, 2000). In another study, Dovidio and Gaertner (2000) found that although 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 ambiguous or uncertain contexts. Considerable evidence demonstrating the operation of covert and unconscious racial and gender biases in a variety of social domains provides reason to suspect that such biases can also influence customer satisfaction judgments.

Our theoretical arguments supporting the influence of racial and gender biases in customer satisfaction judgments are based on the idea that observers (e.g., customers) have preconceived expectations about others depending on whether the person being observed belongs to a high- or low-status demographic group (Berger, Conner, & Fisek, 1974; Berger, Fisek, Norman, & Zelditch, 1977). In the United States, men and whites are considered by most people to be members of a high-status social group relative to women and ethnic minorities (see Ridgeway [1991] for a review). One of the benefits of belonging to a high-status social group is that observers are more likely to make favorable inferences about one's competence, normality, and legitimacy (Aquino & Bommer, 2003; Giannopoulos, Conway, & Mendelson, 2005; Sidanius & Pratto, 1999). In contrast, members of low-status groups are subject to negative stereotypes and attributions concerning their work-related competencies (Fernandez, 1981; O'Leary & Ickovics, 1992).

In view of rating theory (Wherry & Bartlett, 1982), we believe there are at least three reasons to expect customer satisfaction ratings to be susceptible to negative stereotypes and racial and gender biases. First, an important difference between performance ratings made by supervisors and judgments made by customers is that usually customers are afforded the luxury of anonymity. Although supervisors' and customers' ratings are both viewed by organizational administrators, in most cases only supervisors are identifiable. Customer anonymity decreases accountability and the desire to engage in the effortful cognitive processing required to conceal or overcome any biases (Lerner & Tetlock, 1999; Richman, Kiesler, Weisband, & Drasgow, 1999). Moreover, supervisors, but not customers, know that their ratings are part of the employee

record and are used for employee training, feed-back, and advancement decisions (Murphy, 1991). Supervisors can not only be identified, but also must justify their ratings, and as such they are even more motivated to engage in effortful information processing to help them reduce the influence of racial or gender bias and appear, at least superficially, to be objective.

Customer anonymity does not motivate raters to reduce bias, and customer satisfaction questionnaire instructions and items may even facilitate the expression of such biases. Supervisors completing a performance appraisal are frequently reminded of the importance of rating accuracy, and they rate specific behavioral items (Judge & Ferris, 1993; Kane, Bernhardin, Villanova, & Peyrefitte, 1995). Customers are typically asked only for their "opinions" or "attitudes" about employees or organizations, which suggests that they "make a judgment" (Schneider, Ehrhart, Mayer, Saltz, & Niles-Jolly, 2005). Common customer satisfaction items such as "I would recommend this organization to others" and "This organization/employee meets my expectations" do not solicit recall of specific employee or organizational behaviors, and so they may provide more information about the state of mind of the rater than about the actual performance of the ratee or the organization. Such items are also problematic because customers may have higher expectations when rating women and nonwhites (Biernat & Kobrynowicz, 1997; Yarkin, Town, & Wallston, 1982) and so provide lower customer satisfaction judgments for such workers, even if their performance is objectively equivalent to that of their male or white counterparts.

A variety of techniques has been shown to effectively reduce bias in performance appraisal. For example, Roch and O'Sullivan (2003) found that a combination of frame-of-reference (i.e., having raters establish a prototype of good performance) and behavioral observation training (i.e., focusing raters on specific behaviors) increased accuracy in appraisal. Baltes and Parker (2000) found that halo error training (i.e., knowing what factors should not influence ratings) and structured recall memory intervention (i.e., memory enhancement techniques) reduced bias in performance ratings. DeNisi, Robbins, and Cafferty (1989) argued that behavioral diaries aided in the recall and categorization of behavioral events. Customers are not trained in or expected to use these techniques when forming satisfaction judgments.

In sum, customer satisfaction judgments are likely to be highly susceptible to racial and gender biases because customers are usually anonymous, are asked to make summary judgments rather than

to accurately recollect performance-related behaviors, and are untrained in techniques that might help them overcome unconscious biases. Our arguments led us to expect that, in general, employees belonging to low-status demographic groups (women, racial minorities) will receive lower customer satisfaction scores than employees belonging to high-status demographic groups. But even if we found evidence for this difference, it would not demonstrate the operation of bias in customer judgments, because the lower customer satisfaction judgments received by members of lower-status demographic groups might in fact indicate lower levels of true performance. Logically, customer satisfaction judgments should be at least partly influenced by employee objective performance (Wherry & Bartlett, 1982). Consequently, more direct evidence of bias would be demonstrated if behaviors performed by a high-status group member are viewed more favorably by customers than *the same behaviors* performed by someone from a low-status group. Evidence for the plausibility of this hypothesis comes from studies showing that women in leadership roles are rated lower than men in similar roles (Eagly, Makhijani, & Klonsky, 1992) and that ethnic minorities and women are rewarded less than whites and males for exhibiting the same advice-giving or ingratiatory behaviors (Westphal & Stern, 2007). There is also evidence that racial minorities and women who achieve performance equivalent to that of whites and men are judged as having less underlying ability (Biernat & Kobrynowicz, 1997; Yarkin et al., 1982). To examine whether prior findings of biases in judgments of competence based on a target person's group membership generalize to customer judgments, we tested the following hypothesis:

Hypothesis 1. The relationship between employee objective performance and customer satisfaction judgments is attenuated for employees belonging to low-status demographic groups compared to employees belonging to high-status demographic groups.

Hypothesis 1 deals with customer judgments that ask for evaluations about an individual service provider. But we also believe that racial and gender biases can influence evaluations of the service provider's organizational context (e.g., its cleanness or appearance). Our prediction is drawn from the notion that the positive or negative properties of an item or person can "spill over" onto the nearby context or surrounding targets (Rozin, Millman, & Nemeroff, 1986). Our logic is based on the simple idea that evaluations of different aspects of service experience (e.g., the employee, the context) are

connected in a rater's unconscious and conscious belief system (Argo, Dahl, & Morales, 2009; Morales & Fitzsimons, 2007). Theoretical and empirical work on the cognitive structure of attitudes (Anderson, 1981; Wyer & Schrull, 1989) has suggested that evaluation of any person, object, or idea is partly based on evaluations of other persons, objects, or ideas with which the target object is linked. Recent work in marketing has elaborated on this idea (Keller, 2003). Research shows that the evaluation of a product, service, or brand is partly related to the evaluation of the persons who are associated with the product, service, or brand (e.g., the person in an advertisement; the service provider) (Folkes & Patrick, 2003; Morales & Fitzsimons, 2007; Simonin & Ruth, 1998). Accordingly, we expect that customers' evaluations of organizational contexts will be unconsciously connected to their evaluations of highly visible employees. Although such an unconscious connection may be unwarranted, we expect customers' attitudes about low-status employees to be reflected in less favorable evaluations of an organizational context where such employees are highly visible.

It may even be possible for customers to make a conscious connection between employee status and the quality of an organizational context (e.g., Rynes, Heneman, & Schwab, 1980; Rynes & Miller, 1983; Spence, 1973). In line with this thinking, job applicants believe that recruiters' competence and thoroughness signal their organization's overall quality (Rynes, Bretz, & Gerhart, 1991). Customers may judge an organizational context's employment of a low-status person as signaling physical and social inferiority to one employing a high-status person. Combining our spillover and signaling arguments, we expected the mere presence of members of lower-status groups in an organizational environment to lead to less positive customer judgments of the environment:

> *Hypothesis 2. Individuals report lower customer satisfaction judgments of an environment context when a highly visible employee in that environment belongs to a low-status (i.e., female, African American) rather than a high-status (i.e., male, white) demographic group.*

All customers are not so susceptible to racial or gender bias that they will evaluate an employee and environment more negatively simply because the employee belongs to a low-status demographic group. We expected the customers most prone to making these types of judgments to be those who hold more negative preexisting attitudes toward females or racial minorities. Individuals with negative preexisting attitudes toward members of low-status groups quickly associate negative words (e.g., "terrible," "nasty," "evil") with pictures of nonwhite or female faces, and they quickly associate positive words (e.g., "laughter," "glorious," "joy") with pictures of white or male faces (Greenwald & Banaji, 1995). We tested this possibility by examining whether the degree to which individuals have unconscious, negative attitudes toward members of low-status groups moderated the effect proposed in our first two hypotheses:

> *Hypothesis 3a. Low-status employees (i.e., women, African Americans) receive even lower customer satisfaction judgments than equally performing high-status employees (i.e., men, whites) when the judges have negative attitudes toward members of low-status groups.*

> *Hypothesis 3b. Contexts employing low-status employees receive even lower customer satisfaction judgments than contexts employing equally performing high-status employees, when the judges have negative attitudes toward members of low-status groups.*

Thus far we have argued that customer satisfaction judgments can be influenced by perceiving a single employee belonging to a low-status group in a service environment. But in many cases customers interact with a variety of employees during a given customer service encounter. For example, when doing something as simple as buying groceries, customers may observe and interact with deli workers, produce employees, cashiers, and baggers. Employees in each of these positions likely have different demographic characteristics. For this reason, another way to examine the possible influence of bias in the customer judgment process is to see whether the demographic composition of *an organizational unit* influences customer judgments of that unit. Extending our theoretical argument to the organizational level, we expected to observe that the degree to which an organizational unit's employees are members of low-status demographic groups will influence customer satisfaction judgments of that organizational unit in such a way that it will be judged as being of poorer quality than an organizational unit whose employees are mostly from high-status demographic groups. But as we noted when predicting the influence of rater bias on judgments of individual service providers, such a finding might reflect true differences in unit performance. Hence, we make this more precise prediction that parallels Hypothesis 1:

> *Hypothesis 4. The relationship between an organizational unit's objective performance and*

FIGURE 1
Conceptual Model of How Bias Influences Customer Satisfaction Ratings[a]
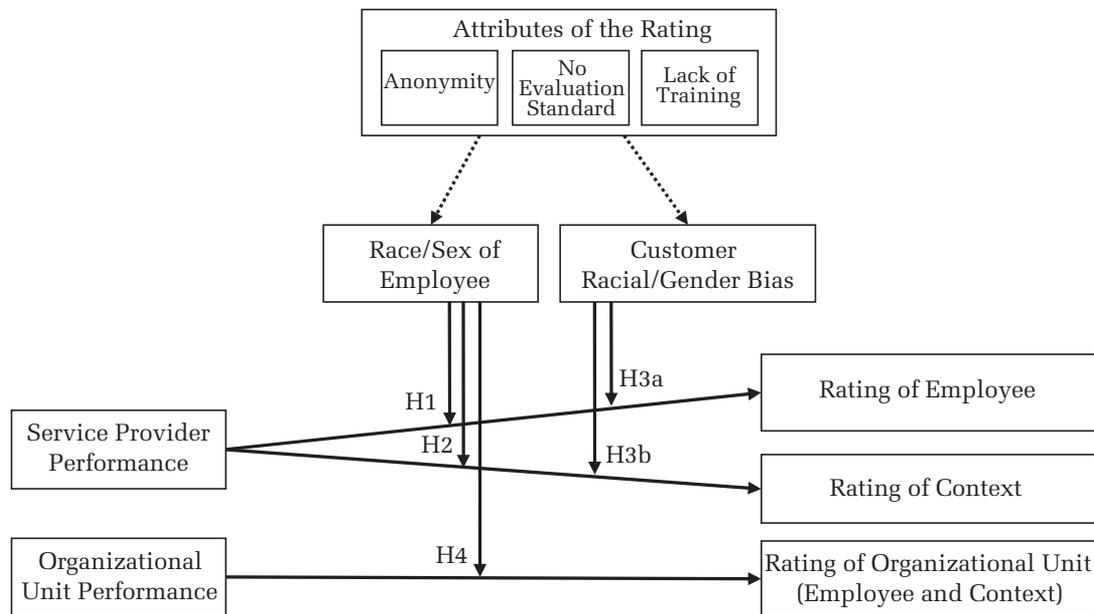
[a] Dotted lines indicate untested relationships. We expected main effects, but our contribution lies with tests of the interactions.

customer satisfaction is attenuated as a function of the percentage of unit employees belonging to low-status demographic groups (i.e., women and minorities).

Our theoretical arguments and hypotheses are summarized in the conceptual model shown in Figure 1.

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

MEDCORP STUDY

Our first sample was drawn from all 113 primary care physicians (i.e., family practitioners) employed by a large health maintenance organization (HMO), hereafter referred to as Medcorp (a pseudonym). Medcorp provides coverage and health care for about 350,000 people in the Pacific Northwest of the United States. Within our sample, 38.4 percent were women, 11.5 percent were ethnic minorities, and all had medical degrees. The 2006 Diversity Report by the Association of American Medical

Colleges reported that 24.5 percent of practicing physicians were women and 12.1 percent were nonwhites.[1]

Measures

Medcorp routinely collected patient satisfaction ratings as well as objective behavioral indicators of physician performance that were assumed to have a direct, positive impact on patient health and well-being. This feature of our data represents a methodological improvement over studies that have only measured employee performance with a single subjective rating and have therefore been unable to determine if the rating was biased (Rotundo & Sackett, 1999). The dependent variable in our study was patient satisfaction with a physician. The independent variables were physician demographic characteristics (race and gender) and three types of patient-centered behaviors.

Customer satisfaction. Medcorp mailed a postcard survey 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

---

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

and returned a total of 12,091 surveys, which constituted a response rate of 52 percent, with each physician rated by an average of 107 patients. Each patient rated only one physician, so the individual ratings were independent. Patients were asked: "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" (1 = "very poor," to 5 = "excellent"). The three items were highly correlated (the average correlation was .93), so the organization combined them to create a composite patient satisfaction variable. These items capture general rather than very specific behaviors (e.g., minutes spent with provider, the number of questions the doctor asked). The organization did not provide us with access to raw patient-level surveys. Instead, it provided us with data indicating what percentage of each physician's patients rated the physician as "excellent." Thus, the range on this measure was from 0 to 100 percent. Physicians who were rated above average on this measure received a 4 percent pay bonus. This measure was collected in the same quarter as all other variables.

Physician race. Medcorp identified each physician's race, and we coded whites as 0 and ethnic minorities as 1. Of the 113 physicians in the sample, 10 were Asian or Pacific Islanders, 2 were black, and 1 was Native American. The percentage of ethnic minority physicians in our sample is consistent with the national average of 12.1 percent.

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

Objectively measured employee performance. With the growing prevalence of 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 behaviors that benefit their patients and the organization's customers (Laine & Davidoff, 1996; Stewart et al., 2000). We used the customer-benefiting behaviors identified by Medcorp as our indicators of objective physician performance.

Medcorp measures customer-benefiting 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 e-mails 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 they spend receiving medical care. For all these metrics, physicians are shown how they compare to both the organizational goal and the organizational average. More productive physicians are able to treat more customer problems per visit, thereby saving customers' time and trips to the doctor. More accessible physicians provide greater convenience to customers, who can simply e-mail their medical questions. Higher-quality physicians are better at preventing costly and deadly health events such as strokes and heart attacks. Physician compensation is tied to each of these customer-benefiting behaviors. Physicians who exceed the 40th percentile are given a bonus.

Physician productivity. The average number of patients seen, medical issues discussed, and medical procedures performed by each doctor in a standardized eight-hour day was recorded by the organization's scheduling software. Medcorp physicians had a great deal of control over the amount of work that they did in a day as they could control the intensity of each visit (e.g., the number of procedures performed and patient health issues addressed). Organizational administrators controlled the number of patients physicians saw each day.

The objective performance assessment we used was the composite of average face-to-face visits and phone visits adjusted by the average intensity of each visit. Intensity was measured by relative value units (RVUs), which physicians coded at the end of each visit using national guidelines. RVUs captured the amount of time involved, the required physical and mental effort, the required judgment and technical skill, and the psychological stress entailed (Hsaio, Braun, Becker, & Thomas, 1988; Hsaio, Braun, Dunn, & Becker, 1988). Physicians checked one of three RVU boxes after seeing each patient. If the patient appointment was a quick check-back or follow-up appointment, physicians checked the first box, which was worth .5 RVUs. If the patient appointment involved at least two patient issues or concerns, but fewer than four, then the physician checked the middle box, indicating 1.0 RVUs for that visit. If the patient appointment involved five or more patient issues, then the physician checked the third box, which indicated 1.5 RVUs. According to quarterly audits by administrators, Medcorp physicians accurately recorded RVUs in 90 percent of patient visits. Coding errors resulting from physicians coding too many or too few RVUs were normally and equally distributed. The raw measure of productivity was standardized

on the basis of physician full-time status and then multiplied by each physician's average visit intensity to obtain the quarterly average RVU-adjusted patient encounters per day.

Physician accessibility. The average daily number of secure e-mails that physicians sent to patients for the quarter was used to measure another type of customer-benefiting behavior. Patients highly value the ability to easily contact their physician. Indeed, a Harris poll has shown that 90 percent of Americans who are online want the ability to e-mail their physicians, and 37 percent are even willing to pay for it (Taylor, 2002). Medcorp patients and physicians could communicate electronically regarding health-related issues through a portal designed exclusively for patient-doctor communication. To use the system, patients logged onto a secure website that provided access to their personal health records, their lab results, and a host of health-related information. Patients could send unlimited e-mails through the portal to any physician they had visited in the prior two years at no cost, and Medcorp physicians were expected to reply to each message within 24 hours. Patients were encouraged to contact their physicians via the system to ask basic health-related questions, to request prescription refills, and to schedule follow-up appointments. Medcorp administrators assigned an equal number of patients to each physician (taking into account patient sickness, age, and gender) and thought that the system saved patients doctor visits, thereby saving them time and money. In general, physicians do not think that e-mail improves the quality of patient care, but rather that it increases convenience for patients (Kleiner, Akers, Burke, & Werner, 2002).

The Medcorp computer server automatically recorded the number of e-mails that each medical professional sent to his or her patients. Medical professionals had a great deal of control over how many e-mails they sent for two reasons: (1) they could try to discourage patients from using the system, and (2) they could choose whether to personally respond to their patients' e-mails. We calculated the number of e-mails physicians sent per day, taking into account the number of full working days that they were in clinic during the study period. To enhance the normality of the variable, we used an inverse transformation and then reflected these values so that higher values represented greater use (Tabachnick & Fidell, 2003).

Physician quality. Every Medcorp primary care physician was responsible for a panel of member-patients. Of the thousands of possible treatments, prescriptions, and procedures that physicians could perform to benefit patients, one of the most important was prescribing 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 in the United States (Willerson & Cohn, 2000), and these drugs prevent cardiovascular events over patients' lifetimes (Gerstein et al., 2000).[2]

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

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

Control Variables

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

Average practice busyness. Patients who had to wait a long time to see their physicians might be

---

[2] We call this variable "quality" because statin and ACE inhibitor prescription rate accuracy are measures of physician quality according to the most influential quality assurance organizations (e.g., HEDIS, NCQA, and IHI).

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

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

Number of patients in panel. Medcorp assigned physicians to care for a particular group (i.e., a 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 might have different expectations about doctor demographic characteristics, so we included the average patient age for each physician's panel in our model.

Average chronic sickness of panel. Sicker patients might be less satisfied, so we controlled for the panel chronic sickness variable calculated by Medcorp, which captured, for example, the per-centage of patients with diabetes and cardiovascular disease.

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

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

Results

Table 1 reports the means, standard deviations, and correlation coefficients for the dependent, independent, and control variables in the Medcorp study. We found no significant differences in our objective measures of performance based on employee race and gender. Our first hypothesis states that the relationship between employee objective performance and customer satisfaction judgments is less positive for employees belonging to low-status demographic groups than for employees belonging to high-status demographic groups. To test this statement, we examined the interactions of the objective measures of employee performance (i.e., quality, productivity, and accessibility) by employee race and gender. We used hierarchical mod-

TABLE 1
Descriptive Statistics, and Correlations, Medcorp Sample[a]

Variable	Mean	s.d.	1	2	3	4	5	6	7	8	9	10	11	12
1. Patient satisfaction	0.51	0.11												
2. Practice busyness	0.66	0.47	−.30											
3. Full-time equivalent	0.80	0.20	−.07	.11										
4. Number of patients in panel	1,749.77	550.63	−.10	.26	.59									
5. Panel age	45.84	4.89	.07	−.07	.05	−.03								
6. Chronic sickness of panel	1.04	0.12	.13	−.12	−.15	−.13	.55							
7. Tenure with Medcorp	14.81	8.51	.20	−.14	.14	.08	.33	−.20						
8. Age	50.34	6.58	.09	−.09	.16	.21	.29	−.05	.69					
9. Nonwhite	0.12	0.32	−.15	.02	.01	−.04	−.14	−.05	−.12	−.03				
10. Female	0.38	0.49	−.06	−.06	−.63	−.44	−.21	.04	−.25	−.31	.12			
11. Productivity	23.00	1.97	.05	.12	.22	.30	−.06	.22	−.25	−.01	−.04	−.15		
12. Quality	−0.01	1.55	.11	.03	.07	.08	.21	.05	.14	.11	.02	.04	−.03	
13. Accessibility to patients	0.16	0.15	.13	−.11	−.18	−.23	.05	−.06	−.04	−.16	−.11	.21	.05	.23

[a] *n* = 113; all correlations larger than .15 are significant at *p* < .05.

erated regression models to do so, centering all variables involved in the interaction terms to minimize multicollinearity between these terms and their individual components (Aiken & West, 1991). We entered all of the control variables in model 1. In model 2 we entered the control variables plus the interactions involving physician gender. In model 3 we entered all the control variables as well as the interactions involving physician race. Finally, in model 4, we entered all control variables and all interaction effects. Table 2 presents the results of this analysis.

The two-way gender by objective performance

TABLE 2
Moderating Effects of Physician Race, Gender, and Objective Performance on Patient Satisfaction with Physician, Medcorp Study[a]

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

[a] $n = 113$. The sample consisted of physicians, 100 whites, 10 Asians or Pacific Islanders, 2 blacks, and 1 Native American.
* $p < .05$
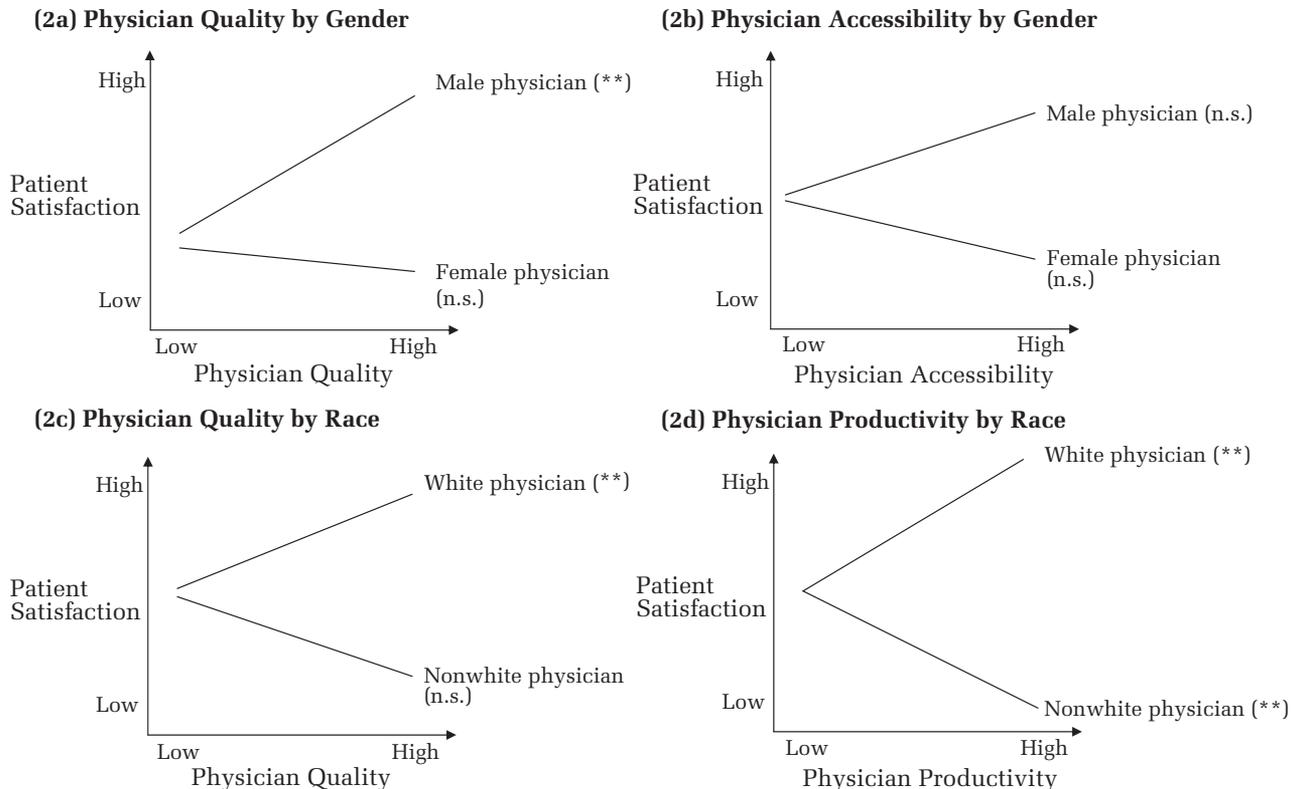** $p < .01$

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

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. The coefficient of the simple slope of quality behaviors on customer satisfaction was significant and positive for male physicians ($b = .32$, $p < .01$), but it was not significant for female physicians ($b = −.01$, n.s.). The coefficient of the simple slope for accessibility behaviors was positive for male physicians ($b = .13$, n.s.) but negative for female physicians ($b = −.17$, n.s.). Although neither of the simple slopes for accessibility behaviors were significantly different from zero, they were significantly different from each other ($p < .05$). By looking at the plots, one can also see that the interaction is a crossover, which shows that the direction of the relationship is the opposite for members of high- versus low-status demographic groups.

The two-way race by objective performance interactions as a set explained a significant amount of incremental variance in the dependent variable ($R^2 = .08$, $p < .05$), providing preliminary support for Hypothesis 1. Inspection of the individual regression weights showed that the physician productivity by race and physician quality by race interactions were significant ($p < .05$). Figure 2 again shows the forms of these interactions. Simple slope analysis revealed that the coefficient of the simple slope of quality on customer satisfaction was significant and positive for white physicians ($b = .29$, $p < .01$) but negative and nonsignificant for nonwhite physicians ($b = −.14$, n.s.). The simple slope of productivity behaviors on customer satisfaction was nonsignificant and positive for white physicians ($b = .15$, n.s.) but significant and negative for nonwhite physicians ($b = −.32$, $p < .01$), which again indicates a crossover interaction effect. Overall, we found support for four of the relationships predicted in Hypothesis 1.

FIGURE 2
Interactive Effects of Physician Objective Performance and Demographic Characteristics on Patient Satisfaction with Physician, Medcorp Study[a]

(2a) Physician Quality by Gender

(2b) Physician Accessibility by Gender

(2c) Physician Quality by Race

(2d) Physician Productivity by Race

[a] Notations in parentheses denote the significance of the simple slopes. "**" = $p < .01$; "n.s." = $p > .10$.

Discussion

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

The observed pattern of relationships indicates that biases against nonwhite and female employees may creep into satisfaction judgments. However, we must also consider this study's shortcomings. First, the relationships between the dependent variable and both panel age and physician age have opposite signs in the correlation table versus the regression table. This pattern suggests that these variables may have somehow influenced our results by suppressing variance in the dependent variable that was irrelevant to its prediction (Tabachnick & Fidell, 2003). However, the correlations are not statistically significant, and our results are substantively unchanged regardless of whether these variables are included in the model, so we believe our results are not due to suppression by those variables. A more serious limitation is that our Medcorp study only included a small percentage of nonwhites. Moreover, many of the nonwhites were Asians rather than African Americans. Biases against African Americans are more negative than those associated with Asians (Song, 2004), and so a study that included African Americans might be better able to detect the influence of such biases on customer satisfaction judgments. We were also not able to control for employees' accents or differences in their language and communication styles, or whether customers felt certain employees had names that sounded nonwhite. It is possible that the biases we observed were the result of some contextual variable such as employee language skill and not of customer prejudices. Finally,

we did not measure whether customer raters had preexisting bias against women and minorities. That is, we had no assessment of the raters' stereotypes or racial/gender biases as potential causes of their ratings. We designed our second study to address the limitations of our Medcorp study. The occupation we chose for our laboratory study was service employees working in a university bookstore, and our raters were college students. We also used an experimental design to control for extraneous variables that might have influenced the results of our Medcorp study.

BOOKCORP STUDY

In our second study, Bookcorp, student raters were asked to observe a video of an employee-customer interaction in a university bookstore, to evaluate the employee's behavior, and to provide satisfaction judgments of the store environment. This study differed from our Medcorp study in a variety of ways. First, we controlled for the job-related behavior of the employees with a scripted interaction and varied only whether the behavior was performed by a male versus a female or a white versus an African American employee. This aspect of the Bookcorp study's design allowed us to reduce variability in employee behavior, thereby providing a better test of whether the same behavior would produce different customer satisfaction judgments depending on the employee's gender or race. Second, we assessed how participants, who were asked to assume the role of customers, not only evaluated the employee (as in our Medcorp study) but also evaluated the organizational context (the bookstore) in which the employee-customer interaction took place. Third, we assessed each participant's implicit bias toward women and non-whites to see if these unconscious attitudes might partly explain gender or racial bias in the ratings.

Sample

Eighty-six students from a major northwestern U.S. public university watched two videos of a bookstore employee interacting with a customer and were asked to evaluate the employee and the bookstore. The bookstore in the video clips was at a large eastern U.S. university, and it is highly unlikely that any of the participants had visited it. The "employees" and "customers" were hired professional actors, and the scripted interaction had been filmed before the bookstore opened in the morning, although our raters taking the customer perspective were not aware of this. Participants were randomly assigned to view either the white

male employee ($n = 33$), white female employee ($n = 21$), or black male employee ($n = 34$). Overall, substantial percentages of our participants were nonwhite (43 percent) or female (38 percent). This heterogeneous sample of raters was representative of the population of people using the book store.

Design

Our design was a mixed factorial, with one between-subjects factor (employee demographic characteristics) and one within-subjects factor (employee-customer interaction). We treated employee demographic characteristics as a between-subjects factor to reduce participants' awareness that they were participating in a race- or gender-related study. We presented all participants with two videos depicting different employee-customer interactions. One video involved the employee ringing up a book and telling the customer that the book's price in the computer was higher than its price on the shelf. The other video showed the same employee being unable to help a customer find a book the customer wanted. Each video was about one minute in length. Each participant saw both videos of the same employee. We randomly assigned the order of the videos' presentation within each experimental condition and found no evidence that order influenced customer ratings. The customer and employee interactions were scripted to ensure that behavior was equivalent across conditions. All videos were filmed in the same bookstore so the store background was equivalent across conditions. The interactions in the video were designed to depict a moderate level of service quality to avoid "floor effects" (nearly everyone would rate performance as very poor) and "ceiling effects" (nearly everyone would rate performance as very good). We reasoned that making employee performance more ambiguous would allow us to more effectively detect the influence of unconscious biases on ratings of service quality and context because it has been shown that the effects of stereotypic biases can be weakened when people are asked to evaluate employees who are performing extremely well or poorly (Dovidio & Gaertner, 2000). We assessed whether the videos did in fact produce ambiguous judgments of employee service quality by asking participants to report their satisfaction with the employee using the customer satisfaction with employee items described in the next section. Results showed that the employee behaviors were perceived as being somewhat negative, as evidenced by satisfaction ratings of the videos averaging 3.27 on a seven-point scale. That this rating was not extremely negative suggests that our videos were

not susceptible to floor effects. After watching each video, participants completed a distraction task (an unrelated questionnaire) to clear short-term memory before they rated satisfaction and context. This procedure simulated the real-world situation in which customers typically do not complete service quality ratings immediately after observing an employee, but only after having engaged in other activities (e.g., dinner, shopping) that can take their attention away from their encounter with the employee.

Dependent Variables

Customer satisfaction with the employee. Raters were asked how satisfied they were with (1) speed of service, (2) quality of service, (3) availability of staff for assistance, and (4) employee responsiveness to customers' issues and concerns (1 = "very poor," 7 = "excellent"). Coefficient alpha for the items was .74. This measure was adapted from an existing customer satisfaction survey we obtained from a large organization. The Appendix gives the customer satisfaction items used in the three studies.

Customer satisfaction with the context. Raters were asked 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 (1 = "very poor," "less than expected," "definitely would not," or "strongly disagree"; 7 = "excellent," "better than expected," "definitely would," or "strongly agree"). This measure (α = .76) was also adapted from an existing customer satisfaction survey we obtained from a large organization.

Predictor Variables

Experimental condition. We had two conditions—one for race and one for sex. The sex condition included participants who viewed the white male or white female employee (1 = "participant viewed two videos of a white female employee," 0 = "participant viewed two videos of a white male employee"). The race condition included participants who viewed either the white or the nonwhite male (1 = "participant viewed two videos of a nonwhite male employee," 0 = "participant viewed two videos of a white male employee").

Participants completed survey questions only after watching both videos.

Implicit bias. To measure raters' racial and gender prejudices, we administered two Implicit Association Tests (IATs) constructed to capture each participant's level of unconscious bias against nonwhites and women (Greenwald, Nosek, & Banaji, 2003). We should note that shortcomings of IAT measurement (Blanton & Jaccard, 2006), such as no absolute zero point and equal intervals, makes the interpretation of scores somewhat unclear. However, we chose to use the IAT as opposed to other types of bias measures (e.g., the modern racism scale [McConahay, 1986]) because it is more difficult for participants to hide prejudices on the IAT than on explicit measures (Nosek, 2005). The gender IAT was administered after the participants saw the videos and made their customer satisfaction judgments, but the race IAT was administered between the videos and the ratings. Prior research indicates no evidence of order effects for the IAT and dependent variables—probably because subjects still respond in socially desirable ways on the explicit measures (Greenwald, Poehlman, Uhlmann, & Banaji, 2009).

Importantly, implicit attitudes appear to be better predictors of behavior than their explicit counterparts, especially when social sensitivity concerns are high (Greenwald et al., 2009). For instance, implicit (but not explicit) attitudes about African Americans have been shown to predict desire to work with an African American partner on an intellectual task (Ashburn-Nardo, Knowles, & Monteith, 2003), and nonverbal actions (eye contact and other "friendly" behaviors) toward African American interaction partners (McConnell & Leibold, 2001). Though the correlation between implicit and explicit attitudes varies across domains (Nosek, 2005), the predictive validity of each suggests that they represent independent processes that explain unique variance in behavioral outcomes (see Greenwald et al. [2009] for a meta-analysis of the predictive validity of the IAT).

Control Variables

We controlled for rater race, gender, and age to account for rater demographic characteristics that might plausibly influence reactions to employee demographic characteristics.

Results

We regressed the employee and organizational context judgments on our controls, predictors, and interaction to determine the degree to which the

TABLE 3
Effects of Employee Race and Gender on Customer Satisfaction with Employee, Bookcorp Study[a]

| | Customer Satisfaction with Employee | | | | | |
| | Female Condition | | | Nonwhite Condition | | |
Variables[b]	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Nonwhite	−.25*	−.26	−.27	−.23*	−.21	−.20
Female	−.15	−.23	−.21	−.15	−.15	−.16
Age	−.09	−.05	−.05	−.22*	−.23*	−.21
IAT score	.05	.15	−.13	−.15	−.14	−.19
Female condition		−.28*	−.26*			
IAT score × female condition			−.06			
Nonwhite condition					−.02	−.01
IAT score × nonwhite condition						−.28**
Adjusted R^2	.03	.08	.08	.08	.08	.14
R^2	.10	.16	.16	.14	.14	.22
ΔR^2 from previous model		.06*	.00		.00	.08**

[a] For the nonwhite experimental condition, 1 = "nonwhite male employee," 0 = "white male employee"; $n = 67$. For the female condition, 0 = "white male employee," 1 = white female employee"; $n = 54$.

[b] "IAT" is the Implicit Association Test.

  * $p < .05$
  ** $p < .01$

customer judgments reflected race and gender bias. Tables 3 and 4 present results of the regression models we used to test Hypotheses 1, 2, and 3.

Hypothesis 1 states that the relationship between employee objective performance and customer satisfaction judgment is less positive for employees belonging to low-status demographic groups than for employees belonging to high-status demographic groups. Recall that the objective performance of each employee was made comparable by using an equivalent behavioral script. Model 2 in Table 3 shows that the raters were significantly less satisfied with women employees than with their equally performing white male counterparts ($\Delta R^2 =$

TABLE 4
Effects of Employee Race and Gender on Customer Satisfaction with Organizational Context, Bookcorp Study[a]

| | Customer Satisfaction with the Context | | | | | |
| | Female Condition | | | Nonwhite Condition | | |
Variables[b]	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Nonwhite	−.03	−.03	−.07	−.16	−.08	−.08
Female	.09	−.03	.09	.16	.01	.01
Age	−.26	−.19	−.14	−.17	−.04	−.04
IAT score	−.12	.03	−.04	.13	.09	.02
Female condition		−.45**	−.38**			
IAT score × female condition			−.23*			
Nonwhite condition					−.44***	−.46***
IAT score × nonwhite condition						−.18*
Adjusted R^2	.00	.15	.17	.00	.18	.20
R^2	.07	.24	.29	.07	.22	.26
ΔR^2 from previous model		.17**	.04**		.15***	.04*

[a] For the nonwhite experimental condition, 1 = "nonwhite male employee"; 0 = "white male employee"; $n = 67$. For the female condition, 0 = "white male employee," 1 = white female employee"; $n = 54$.

[b] "IAT" is the Implicit Association Test.

  * $p < .05$
  ** $p < .01$
  *** $p < .001$

.06; $b = -.28$, $p < .05$). However, we did not find evidence of bias in customer satisfaction judgments of the nonwhite employee ($b = -.02$; n.s.). Thus, Hypothesis 1 received partial support.

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

Hypothesis 3a suggests that people report even lower customer satisfaction judgments of an employee belonging to a low-status demographic group when they have negative implicit attitudes toward that group. Model 3 of Table 3 shows that the coefficient for the interaction of IAT score by

nonwhite condition is significant and in the expected direction for customer satisfaction with the employee ($\Delta R^2 = .08$; $b = -.28$, $p < .01$). To gain more insight into this effect, we plotted the interaction and analyzed the simple slopes, as shown in Figure 3. Individuals with high levels of implicit bias (+1 s.d.) were significantly more likely to report lower satisfaction with the nonwhite male's performance than with the white male's ($p < .01$). However, the coefficient for the interaction of IAT score by gender condition was not significant.

As for Hypothesis 3b, model 3 of Table 4 shows that the coefficient for the interaction of IAT score and race is significant and in the expected direction for customer satisfaction with the rated context ($\Delta R^2 = .04$; $b = -.18$, $p < .05$). We plotted the interactions and conducted a simple slope analysis (see Figure 3). Customer IAT score (+1 s.d.) was positively related to customer satisfaction with the context when customers were observing a white male employee ($b = .33$, $p < .01$) but was negatively related to customer satisfaction with the context when customers were observing a nonwhite male employee ($b = -.21$, $p < .05$). The coefficient

FIGURE 3
Interactive Effects of Employee Demographic Characteristics and Customer Implicit Association Test Score on Customer Satisfaction with Employee and Store Context, Bookcorp Study[a]

(3a) Employee Satisfaction by Race

(3b) Context Satisfaction by Race

(3c) Context Satisfaction by Gender
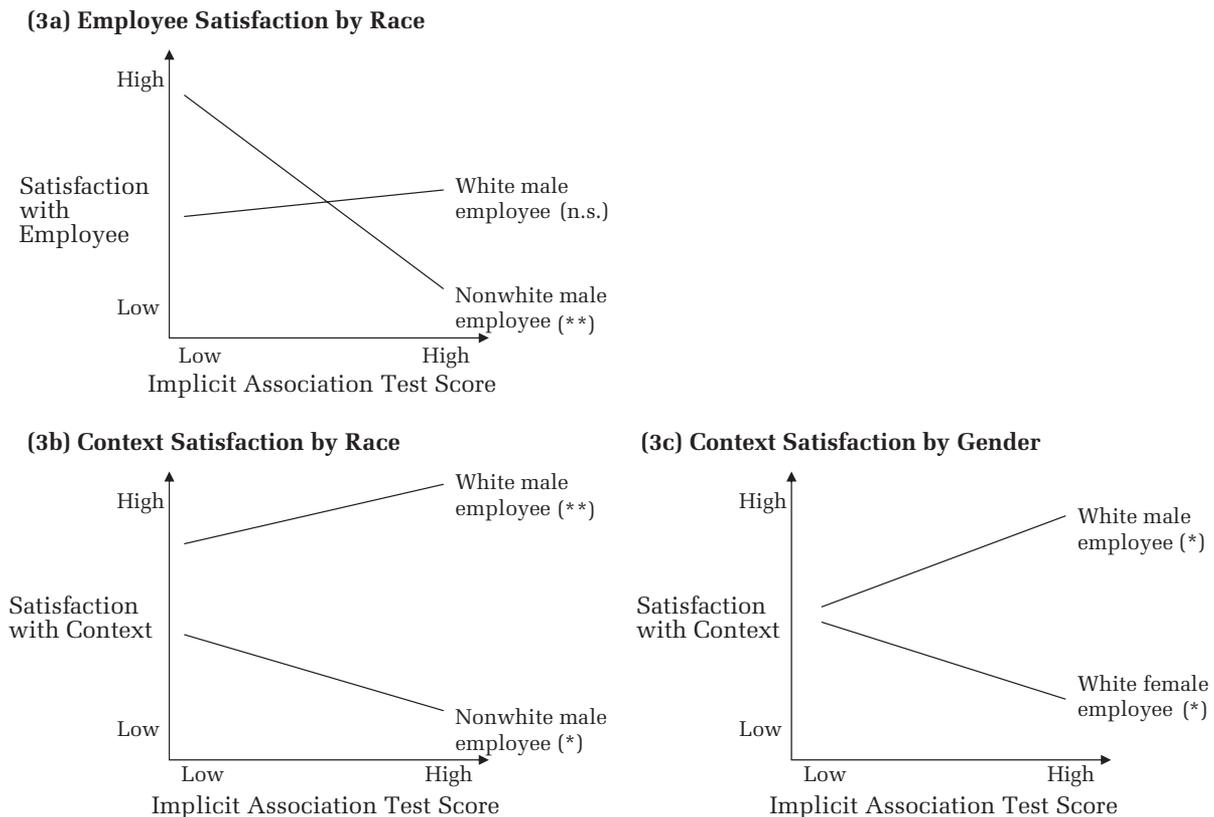
[a] Notations in parentheses refer to the equation represented by the slope. "**" = $p < .01$; "*" = $p < .05$; "n.s." = $p > .10$.

for the interaction term IAT score by gender condition was significant and in the expected direction for customer satisfaction with the context ($\Delta R^2 = .04$; $b = -.23$, $p < .05$). Customer IAT score ($+1$ s.d.) was positively related to customer satisfaction with the context when customers were observing a white male employee ($b = .23$, $p < .05$) but was negatively related to customer satisfaction with the context when customers were observing a white female employee ($b = -.21$, $p < .05$). Overall, we found three significant coefficients supporting Hypothesis 3. These results and plots suggest that judgments of both employees and organizational contexts are vulnerable to unconscious biases.

Discussion

We found that student participants taking a customer perspective rated employees and an organizational context as worse when observing the performance of a low-status employee, and this was especially true if the raters held implicit biases about that low-status group. The results for Hypothesis 1 were not as strong as we anticipated, suggesting that it may be possible to minimize biases by changing the setting in which rating takes place. The laboratory context was less anonymous than a typical customer satisfaction questionnaire setting, which may have weakened the influence of bias on customer satisfaction judgments of employees. Although we told the participants that their responses were anonymous, they may have felt scrutinized because they provided their judgments when an experimenter was present and wrote their names on a separate sign-in sheet to receive participation credit for a class. Additionally, administration of the IAT for race prior to the customer satisfaction ratings may have alerted participants that they were in a race and gender study, which may have weakened the Bookcorp study results for race. We observed, though, that the effects of bias on judgments of the organizational context were still quite strong, which makes sense to us because participants may have been unaware of and therefore unable to suppress biases that spilled over onto their judgments of the organizational context.

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

GOLFCORP STUDY

Our sample was drawn from a large country club organization, hereafter referred to as Golfcorp. Golfcorp has 66 country clubs across the United States and roughly 70,000 customer-members, and it employs approximately 8,000 people. Our sample consisted of all 66 Golfcorp country clubs. In our sample, 31.4 percent of employees were women, 18.1 percent were Latino, 6.7 percent were African American, and 1.7 percent were Asian American or Native American.

Measures

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

Customer satisfaction with facility. Like many organizations, Golfcorp measures customer satisfaction with a quarterly survey, which is mailed to a percentage of each facility's customers. An average of 63.8 customers rated each facility. The average response rate per facility was 27.3 percent (an average of 234 surveyed customers per facility). The marketing company hired to do the customer survey randomly sampled each facility's customers each quarter until they got either 20 respondents or 3 percent of the total customer base (whichever was larger). The items used for this measure reflected a focus on the facility context (quality of its clubhouse and golf course) and overall ratings of the facility; it was thus similar to what was used in the Bookcorp study. Customers rated each of these items ($\alpha = .81$): "How would you rate the following aspects of your club . . . " (1) "maintenance of grounds/appearance of clubhouse," (2) "locker rooms and restrooms," (3) "quality of greens," (4) "condition of course," (5) "pace of play," (6) "condition of practice facilities," (7) "ability to obtain desired tee times" (items 1–7 were rated as 1 = "very poor," 5 = "very good"); (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"). In our analyses, we lagged

this measure six months after the independent variables to more conclusively show that employee demographic characteristics and objectively measured performance influence customer ratings, rather than the other way around.

Percent nonwhite. Golfcorp identified the percentages of white and nonwhite employees in each facility. Across the 66 clubs in the sample, 26.4 percent of employees were nonwhite. According to the U.S. Census Bureau's 2000 census, the percentage of ethnic minority employees in our sample was consistent with the national average of 28 percent of the U.S. population.

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

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

Facility productivity. Facility productivity was calculated by Golfcorp's central accounting office for the calendar year ending six months before the dependent variable was collected. This variable was each club's annual profits divided by its average number of employees in that year. The number of employees at each club was centrally controlled, with clubs with more members allotted proportionally more employees by the central office. Therefore, facility productivity was determined by the employees' effectiveness at creating value.

Facility quality. Over time, the condition of golf courses and clubhouses deteriorates, and they need to be rebuilt or refurbished. Golfcorp assessed the quality of the course and clubhouse of each club to ensure that customers were receiving a high standard of service. For the courses, the rating scale was 1, "more than 40% of course is crab grass or dead spots," to 5, "less than 5% of course is crab grass or dead spots." For the clubhouses, 1 was "built or refurbished more than 15 years ago" and 5 was "built or refurbished 2 years ago or less." The overall facility attribute quality variable was the composite of these scales. The two component variables approached normality and were added together.

Control variables. Customers may be more satisfied with larger facilities because they offer more amenities, and they may be more satisfied with facilities that employ large percentages of young or temporary employees, who may be more energetic. Therefore we controlled for *facility size*, *average employee age*, and *percent temporary employees*. Customers may be more satisfied if they have been members for a long time and have not quit, if they are men, or if they are older. Therefore, we also controlled for *average customer tenure* (months), *percent male customers*, and *average customer age*.

Results

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

Hypothesis 4 states that the association between an organizational unit's objective performance and customer satisfaction will be attenuated for organizational units that employ higher percentages of employees belonging to low-status demographic groups compared to units that employ higher percentages of employees belonging to high-status demographic groups. The two-way gender by objective performance interactions as a set explained a significant amount of incremental variance in the dependent variable ($\Delta R^2 = .04$, $p < .05$) providing some further support for this hypothesis. Inspec-

TABLE 5
Descriptive Statistics and Correlations, Golfcorp Study[a]

Variable	Mean	s.d.	1	2	3	4	5	6	7	8	9	10
1. Customer satisfaction with facility	3.90	0.22										
2. Size	129.05	157.51	.16									
3. Average employee age	39.20	4.04	−.18	.09								
4. Percent temporary employees	0.04	0.08	.15	−.07	.16							
5. Average customer tenure	60.66	22.10	−.15	.14	.15	.01						
6. Percent male customers	0.56	0.07	.00	−.03	−.13	.06	−.03					
7. Average customer age	54.23	8.40	−.14	.17	.34	.01	.42	−.16				
8. Percent nonwhite employees	0.26	0.18	.08	−.12	.11	.20	−.28	.15	.08			
9. Percent female employees	0.31	0.12	−.07	−.21	.07	.05	−.07	.20	−.06	−.08		
10. Facility quality	3.32	1.04	.08	.16	−.14	.16	−.04	.16	−.03	.10	−.08	
11. Facility productivity	14,614.26	7,703.06	.13	.14	.12	.16	.12	.13	.16	.14	.09	.15

[a] $n = 66$; all correlations greater than .21 are significant at $p < .05$.

tion of the individual regression weights in the full model showed that the facility quality by gender interaction and the facility productivity by gender interaction were significant ($p < .05$). We probed the pattern of the interactions by examining the simple slopes of the objective performance measures for facilities with high and low percentages of female employees (Aiken & West, 1991).

As in the Medcorp study plots, a stronger positive relationship between objective performance and customer satisfaction emerged for facilities that had low percentages of female employees than for facilities that had high percentages of female employees. We do not report the plots owing to space limitations, but they are available from the first author. Facility quality was significantly more pos-

TABLE 6
Interactive Influence of Percent Nonwhite Employees, Percent Female Employees, and Objective Facility Performance on Customer Satisfaction with Facility, Golfcorp Study[a]

Variables	Customer Satisfaction with Facility			
	Model 1	Model 2	Model 3	Model 4
Controls				
Size	.21	.27*	.27*	.30**
Average employee age	−.21	−.24*	−.24*	−.19
Percent temporary employees	.18	.16	.16	.18
Average customer tenure	−.12	−.06	−.06	−.06
Percent male customers	−.06	−.08	−.08	−.11
Average customer age	−.09	−.13	−.13	−.10
Facility quality	−.04	−.01	−.01	−.03
Facility productivity	.17	.01	.01	−.05
Percent nonwhite employees	−.05	−.15	−.15	−.19
Percent female employees	.04	−.07	−.07	−.14
Interactions				
Percent female × quality		−.25*		−.23*
Percent female × productivity		−.22		−.31*
Percent nonwhite × quality			−.25*	−.25*
Percent nonwhite × productivity			−.34**	−.49**
Adjusted R^2	.00	.03	.12	.20
R^2	.13	.17	.28	.37
ΔR^2 from model 1		.04*	.15**	.24***

[a] $n = 66$ country clubs.
* $p < .05$
** $p < .01$
*** $p < .001$

itively associated with customer satisfaction for facilities with low percentages of female employees than for facilities with high percentages of female employees. Likewise, facility productivity was more positively associated with customer satisfaction for facilities with low percentages of nonwhite employees than for facilities with high percentages of nonwhite employees. Again, even though the simple slopes were not significantly different from zero, the significant regression coefficient in the full model demonstrates that they were significantly different from each other. By looking at the plots, one can see that the interaction is a crossover, which shows that the direction of the relationship is the opposite, and statistically significant, for members of high- versus low-status demographic groups. These results support Hypothesis 4 for both of the gender by objective performance relationships.

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

Discussion

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

GENERAL DISCUSSION

We set out to determine if and how customer satisfaction judgments are influenced by racial and gender biases. In three samples, we found converging evidence that customer satisfaction judgments are susceptible to systematic and predictable racial and gender biases. Customers tended to be less satisfied with the services provided by women and nonwhite employees rather than by men and white employees, even when objective indicators of performance were controlled for (Medcorp study). We also found that these biases operated on judgments about the service context when a third-party evaluator observed an employee interacting with a customer (Bookcorp study), particularly if the observer had implicit negative attitudes about women or minorities. Finally, we found that evaluations of an organizational unit as a whole were negatively associated with the presence of nonwhite and female employees (Golfcorp study). It is worth noting that we found evidence for the operation of racial biases regardless of whether the nonwhite employees were predominantly Asian (Medcorp study), African American (Bookcorp study), or Latino (Golfcorp study).

The consistency of our results across three different samples and methodologies testifies to the robustness and generality of the systematic biases we observed and to the internal validity of our theoretical model. The pattern of these biases may help explain the persistence of demographic inequalities in organizations. To cite just a few examples, per 2006 U.S. Census Bureau data women and nonwhites make 25 percent less than their male and white counterparts in equivalent jobs (http://www.census.gov/hhes/www/income/histinc/p38N.html), women and nonwhites are twice as likely as white men to be unemployed and underemployed (National Institute for Occupational Safety and Health, 2002), and women and ethnic minorities are not well represented among the ranks of highly paid managers and professionals in U.S. corporations and in prestigious occupations like law and medicine (e.g., Baldi & McBrier, 1997; Eagly & Karau, 2002; Wilson, Sakura-Lemessy, & West, 1999). Economists have often been perturbed by these demographic inequalities because orthodox economic theory would predict that some of these inequalities should be erased when employers compete for women and nonwhite applicants, whose wages are 25 percent less costly than their white and male counterparts (*Economist*, 2008). Our results suggest that the evidence available to organizational decision makers is that customers tend to be less satisfied with nonwhite and female employees; however, without the benefit of the research conducted here, decision makers would not be able to determine that these lower satisfaction ratings are attributable at least in part to customer biases. Therefore, decision makers might rationally choose, given their limited information, to preferentially select white and male employees, as their data are likely to

indicate that such personnel are better performers. Evidence from highly publicized lawsuits (e.g., Shoney's Restaurant; Abercrombie & Fitch) suggests that managers are keenly aware of the fact that some customers prefer white and male employees. Executives in these cases admitted to deliberately favoring white employees in hiring and promotion decisions to enhance customer satisfaction and organizational profitability (Brief et al., 2000).

One could argue that when customers view employees or units as performing less well because of employee demographic characteristics such as race or gender, these employees and units *should* receive fewer rewards, bonuses, and promotional opportunities. But it is important to note that the customer judgments in our studies were inconsistent with other, objective indicators of performance. In other words, nonwhites and women may have behaved the same way as their white and male counterparts in trying to provide satisfactory customer service, but if compensation and other organizational benefits are linked to customer satisfaction ratings, then they may not be rewarded similarly for identical behavior, which would violate the principle of equity that most business enterprises claim to follow.

Our results suggest that if customer evaluations become widely and uncritically used to determine pay and promotion opportunities, the job outcomes of women and ethnic minorities could be adversely impacted. For example, consider what would happen if managers noticed which employees routinely received the highest customer satisfaction scores and used this information to make promotion decisions or if university administrators relied heavily on student ratings of teaching effectiveness to influence promotion and tenure decisions. At higher levels of an organization, executives might examine which of their organizational units achieved the highest levels of customer satisfaction and promote those managers further up the organizational hierarchy. Our data suggest that one possible consequence of these decisions is that whites and men will be much more likely than their nonwhite or women counterparts to receive favorable customer satisfaction judgments, which should accelerate their journeys up the organizational ladder. Likewise, managers who purposely stock their organizational units with whites and men are likely to have more career success than managers who do not.

Our finding that customer biases can spill over onto the surrounding organizational context contributes to the literature on contamination and signaling and also illustrates the subtle operation of racial and gender bias. Marketing researchers have shown how observable customer characteristics such as physical attractiveness can influence other customers' desire to purchase a product (Argo et al., 2009). To our knowledge, ours is the first study to show how an observable characteristic of an employee, such as race or gender, can influence customer perceptions of an organization's contextual quality. This finding may provide insight into the phenomenon known as "white flight," wherein whites move out of a neighborhood once a critical mass of nonwhites has moved in (Gladwell, 2000; Kruse, 2005). In an organizational setting, a similar phenomenon may operate in which customers link their conscious or unconscious negative attitudes toward members of low-status groups to employees who belong to these groups. In turn, these associations "contaminate" customer perceptions of an organizational context. This process of contextual spillover may partly explain why managers have often been reluctant to pursue diversity, despite the known performance advantages of having a diverse workforce (Joshi, Liao, & Jackson, 2006; O'Reilly, Williams, & Barsade, 1997). It may be that managers are aware that diversity has hidden costs because it increases the possibility of "customer flight" to an organization that has fewer employees who belong to low-status demographic groups. Future studies using customer satisfaction as an outcome variable should take into account the demographic make-up of organizations' employees as well as objectively measured organizational characteristics.

Our findings cast doubt on the ability of customers to accurately perceive the quality of customer service organizations. The theoretical attributes we suggested as possible causes of customer bias were meant to explain why these biases occur, but like any useful theory they also suggest potential remedies. Our theorizing suggests that racial and gender biases in customer satisfaction judgments may be reduced by (1) making customer ratings less anonymous, (2) changing the standards customers use to make their ratings so they emphasize behavior rather than subjective judgments, and (3) introducing customer debiasing education or training into the evaluation process. Identifiability (i.e., nonanonymity) may provide an especially strong debiasing effect in service contexts characterized by repeated interactions (e.g., the doctor-patient relationship) because customers would want to make sure their ratings do not jeopardize the quality of future received service. Behaviorally anchored rating scales may be especially helpful for removing bias in customer expectations, because individuals often hold nonwhites and women to higher standards (Biernat & Kobrynowicz, 1997). Organizations could consider only accepting cus-

tomer satisfaction surveys from customers who have completed debiasing training, although such a move would be logistically and perhaps scientifically problematic (owing to potential selection bias).

In addition to addressing factors that cause bias in customer ratings, organizations can take practical steps to minimize the potential adverse impact of customer biases on nonwhite and female employees' careers. For example, organizations might consider only using satisfaction surveys from frequent customers to ensure that raters have had sufficient exposure to targets. Organizations could also ask for customer feedback during service encounters so that customers will be most likely to be paying attention and less likely to rely on information that is subject to memory bias when judging their experience. Organizations might also want to let customers know that the data will be used to make career progression decisions so that they are more motivated to judge responsibly. Organizations could also insert bias-sensitive questions into customer satisfaction judgments so that responses from potentially biased customers could be given less weight or discarded. However, we urge caution if organizations choose to remove outlier ratings, because this tactic may actually decrease rating accuracy if most judges are aware of their biases and therefore tend to overcorrect their judgments (Zitzewitz, 2006). Alternatively, organizations may be able to statistically correct for bias when calculating customer satisfaction judgments. Finally, using different survey formats for customer rating scales might also be helpful for circumventing rater biases; possibilities are forced choice, behaviorally anchored rating scales (citing specific valued behaviors), and unweighted and weighted checklists. Organizations should consider the trade-offs these formats entail and choose the one that is most likely to reduce the effects of customer judgment biases on the career prospects of those who are the most vulnerable targets.

Limitations

We believe our findings provide strong, consistent support for our theoretical predictions. However, like all research, ours has its share of limitations. First, role congruence may have been an issue in our Medcorp study, as patients may expect their physicians to be white and male and therefore judge nonwhite or women doctors more harshly. However, there is not much evidence to suggest that they do expect their doctor to be a white male. Indeed, patients prefer their doctor to look like themselves, in that women prefer women doctors,

and nonwhite patients prefer nonwhite doctors (Chen, Fryer, Phillips, Wilson, & Pathman, 2005; Cooper-Patrick, Gallo, & Gonzales, 1999). Likewise, role congruence should not have been an issue in our Bookcorp study, so the consistent results found for these two samples provide us with some confidence that role congruence alone is not responsible for our findings. Still, future research would be well served to test our hypotheses in a variety of samples as well as to see whether other observable demographic characteristics, such as age, might influence people's career progress (see Wang, Adams, Beehr, & Shultz, 2009) as a result of the mechanisms we described.

Another potential issue is rater-target congruence. Although including patient gender and race in the analysis slightly strengthened our Medcorp study results, we did not control for these variables in that study because of multicollinearity (that is, the respective correlations between customer gender and race and physician gender and race were greater than .90). In our other two studies, we found no evidence that rater-target congruence influenced customer satisfaction judgments. Indeed, in post hoc analyses, the interactions of customer race by employee race and customer gender by employee gender were not significant. To maintain compliance with the rule of thumb that there should be five cases per variable (Tabachnick & Fidell, 2003), we do not report these post hoc analyses.

Another potential limitation is that unobserved variables may be responsible for our results. In our Golfcorp study, employee demographic characteristics may have masked the facility's strategy. Facility executives pursuing a low-cost strategy may have hired a large number of women and nonwhite employees, whereas those pursuing a premium-pricing strategy may have hired a large number of whites and men. We ran some post hoc analyses to test this idea and found no supporting evidence. Specifically, we ran the interactions of several variables measuring club strategy (i.e., services offered, turnover rate) by objective performance, and although our four race and gender interactions remained significant, none of the additional interactions were significant (we do not report these analyses to maintain an adequate case-to-variable ratio). Relatedly, we also did not test or report whether nonwhite women faced a double jeopardy (Berdahl & Moore, 2006) in terms of customer satisfaction to maintain compliance with the five-to-one rule of thumb (that is, we were not able to test the three-way interactions involving employee objective performance, employee race, and employee gender), but future research should do so. Finally, we did not test whether the Implicit Association

Test score influenced ratings of organizational units because we found that it influenced subcomponents of the organizational unit ratings—that is, the employee and context ratings.

We should also mention that our method of testing for possible bias in performance evaluations was a significant improvement over past studies. First, we used an objective performance standard so that we could compare subjective judgments with this standard and therefore determine whether race and gender might influence the customer judgments of performance. Second, our subjective judgments were based on aggregated judgments from a large number of customers rather than the judgments of a single supervisor. This is important because the large number of raters provided a highly reliable subjective performance rating for each individual, context, or organization. The IAT is new, and its predictive validity is relatively untested (Blanton & Jaccard, 2006), so another contribution our study offers is the first demonstration of the IAT's predictive validity in a management journal. Finally, we controlled for several variables that could provide alternative explanations for our results, such as average employee age and average customer tenure with an organization.

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

Conclusions

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

If these results are replicated and generalizable, they have significant implications for organizational practice. If managers are serious about the fair treatment of their employees and the promotion of diversity, they need to treat customer ratings differently. More specifically, the rating process can be changed by increasing information, responsibility, or training for raters and by changing how customer ratings are used. In the latter context, organizations can perhaps measure and discount such biases or statistically adjust the ratings to remove the bias. Without such actions, given the increasing dependence on customer ratings, society is not only likely to maintain existing levels of inequitable compensation and advancement for women and minorities, but also likely to increase these inequities. This outcome is unacceptable in a society that is committed legally, morally, and socially to fair treatment for all in the workplace.

REFERENCES

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

Anderson, E. W., Fornell, C., & Mazvancheryl, S. K. 2004. Customer satisfaction and shareholder value. **Journal of Marketing,** 68(4): 172–185.

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

Anderson, N. H. 1981. **Foundations of information integration theory.** New York. Academic Press.

Argo, J. J., Dahl, D. W., & Morales, A. C. 2009. Positive consumer contamination: Responses to attractive others in a retail context. **Journal of Marketing Research,** 45: 690–701.

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 mistreat-

ment: How victim status moderates the relationship between organizational citizenship behavior and workplace victimization. *Organization Science,* 4: 274–285.

Baldi, S., & McBrier, D. B. 1997. Do the determinants of promotion differ for blacks and whites? Evidence from the U.S. labor market. *Work and Occupations,* 24: 478–497.

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

Berdahl, J. L., & Moore, C. 2006. Workplace harassment: Double-jeopardy for minority women. *Journal of Applied Psychology,* 91: 426–436.

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.

Blanton, H., & Jaccard, J. 2006. Arbitrary metrics in psychology. *American Psychologist,* 61: 27–41.

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: Jossey-Bass.

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

Chatman, J. A., & Flynn, F. J. 2005. Full-cycle micro-organizational behavior research. *Organization Science,* 16: 434–447.

Chen, F. M., Fryer, G. E., Jr., Phillips, R. L., Jr., Wilson, E., & Pathman, D. E. 2005. patients' beliefs about racism, preferences for physician race, and satisfaction with care. *Annals of Family Medicine,* 3: 138–143.

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

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

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

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

Davis, W. E., & Davis, D. R. 1999. The university presidency: Do evaluations make a difference? *Journal of Personnel Evaluation in Education,* 13: 119–140.

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

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

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

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

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

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

Economist. 2008. Black America: Nearer to overcoming. May 8: 34–36.

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. *Monthly Labor Review,* 130(November): 53–85.

Folkes, V. S., & Patrick, V. M. 2003. The positivity effect in perceptions of services: Seen one, seen them all? *Journal of Consumer Research,* 30: 125–137.

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

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

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

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

Gladwell, M. 2000. *The tipping point: How little things can make a big difference.* Little, Brown.

Goldin, C., & Rouse, C. 2000. Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American Economic Review,* 90: 715–741.

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

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

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

Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. 2009. Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology:* 97: 17–41.

Gruca, T. S., & Rego, L. L. 2005. Customer satisfaction, cash flow, and shareholder value. *Journal of Marketing,* 69(3): 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.

Hsiao, 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.

Hsiao, W. C., Braun, R., Dunn, D., & Becker, E. R. 1988b.

Resource-based relative values: An overview. *Journal of the American Medical Association,* 260: 2347–2353.

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

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

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

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

Kane, J. S., Bernardin, H. J., Villanova, P., & Peyrefitte, J. 1995. Stability of rater leniency: Three studies. *Academy of Management Journal,* 38: 1036–1051.

Keller, K. L. 2003. *Strategic brand management.* Upper Saddle River, NJ: Prentice Hall.

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

Kruse, K. M. 2005. *White flight: Atlanta and the making of modern conservatism.* Princeton, NJ: Princeton University Press.

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

Landy, F. J., Shankster, L. J., & Kohler, S. S. 1994. Personnel-selection and placement. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology,* vol. 45: 261–296. Palo Alto, CA: Annual Reviews.

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.

Lerner, J. S., & Tetlock, P. E. 1999. Accounting for the effects of accountability. *Psychological Bulletin,* 125: 255–275.

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.

McConahay, J. B. 1986. Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism:* 91–126. New York: Academic.

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.

Mercer Consulting Group. 2007.

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

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

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

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

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

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

O'Reilly, C. A., Williams, K. Y., & Barsade, S. G. 1998. Group demography and innovation: Does diversity help? In D. Gruenfeld, B. Mannix, & M. Neale (Eds.), *Research on managing groups and teams:* 183–207. Stamford, CT: JAI Press.

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

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

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

Ridgeway, C. L. 1991. The social construction of status value: Gender and other nominal characteristics. *Social Forces,* 70: 367–386.

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

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

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

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

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

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

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

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

Sidanius, J., & Pratto, F. 1999. *Social dominance: An intergroup theory of social hierarchy and oppression.* Cambridge, U. K.: Cambridge University Press.

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

Simonin, B. L., & Ruth, J. A. 1998. Is a company known by the company it keeps? Assessing the spillover effects of brand alliances on consumer brand attitudes. *Journal of Marketing Research,* 35: 30–42.

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

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

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

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

Tabachnick, B. G., & Fidell, L. S. 2003. *Using multivariate statistics* (5th ed.). Boston: Allyn & 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(8): 1–3.

Wang, M., Adams, G. A., Beehr, T. A., & Shultz, K. S. 2009. Career issues at the end of one's career: Bridge employment and retirement. In S. G. Baugh & S. E.

Sullivan (Eds.), *Maintaining focus, energy, and options through the life span:* 135–162. Charlotte, NC: Information Age Publishing.

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

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

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

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

Wilson, G., Sakura-Lemessy, I., & West, J. P. 1999. Reaching the top: Racial differences in mobility paths to upper-tier occupations. *Work and Occupations,* 26: 165–186.

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

Wyer, R. S., Jr., & Schrull, T. K. 1989. Person memory and judgment. *Psychological Review,* 96: 58–83.

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.

Zitzewitz, E. 2006. Nationalism in winter sports judging and its lessons for organizational decision making. *Journal of Economics & Management Strategy,* 15: 67–99

APPENDIX

Customer Satisfaction Items

Satisfaction with Employee

Medcorp
   "How would you rate the following attributes of your provider?" (1 = "very poor"; 5 = "excellent"):

1. Attention provider paid
2. Thoroughness and competence of provider
3. Ability to ask questions of this provider

Bookcorp
   "How would you rate the following?" (1 = "very poor"; 7 = "excellent"):

1. Speed of service
2. Quality of service
3. Availability of staff for assistance
4. Employee responsiveness to customers' issues and concerns

Satisfaction with Context

Bookcorp
   "How would you rate the following aspects of the bookstore?"

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

Golfcorp (Satisfaction with 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")

———————⋀⋀⋀———————

David R. Hekman (hekman@uwm.edu) is an assistant professor of management at the Lubar School of Business, University of Wisconsin–Milwaukee. He earned his Ph.D. in management from the University of Washington. He is interested in improving organizational health by minimizing organizational problems such as weak employee attachment, contagious harmful behaviors, and persistent workplace inequality.

Karl Aquino (karl.aquino@sauder.ubc.ca) is the Richard Poon Professor of Organizations and Society at the Sauder School of Business, University of British Columbia. His research interests are moral behavior, workplace victimization, revenge and forgiveness, and status and power dynamics in organizations. He received his Ph.D. in organizational behavior from Northwestern University.

Bradley P. Owens (bpowens@bus.umich.edu) is a postdoctoral fellow at the Center for Positive Organizational Scholarship at the University of Michigan. He earned his Ph.D. in management from the University of Washington. His primary research interests include humility, leadership, race and gender issues, and team dynamics.

Terence R. Mitchell (trm@u.washington.edu) received his Ph.D. from the University of Illinois in organizational psychology. He is the Carlson Professor of Management at the University of Washington. Research interests include motivation, leadership, turnover, and decision making.

Pauline Schilpzand (pauline.schilpzand@usma.edu) is an assistant professor of character development at the Army Center of Excellence for the Professional Military Ethic at West Point. She earned her Ph.D. in management from the Warrington College of Business Administration at the University of Florida. Her primary research interests include behavioral ethics, leadership, and the selection process.

Keith Leavitt (keith.leavitt@usma.edu) is an assistant professor of character development at the Army Center of Excellence for the Professional Military Ethic at West Point. He earned his Ph.D. in management from the Foster School of Business at the University of Washington. His primary research interests include behavioral ethics, implicit social cognition, research methods, and organizational citizenship.

———————⋀⋀⋀———————