2011

Is the Customer Always Right? The Potential for Racial Bias in Customer Evaluations of Employee Performance

Michael Lynn
Cornell University, wml3@cornell.edu

Michael C. Sturman
Cornell University, mcs5@cornell.edu

Follow this and additional works at: https://scholarship.sha.cornell.edu/articles

Part of the Applied Behavior Analysis Commons, Food and Beverage Management Commons, and the Race and Ethnicity Commons

Recommended Citation

This Article or Chapter is brought to you for free and open access by the School of Hotel Administration Collection at The Scholarly Commons. It has been accepted for inclusion in Articles and Chapters by an authorized administrator of The Scholarly Commons. For more information, please contact hotellibrary@cornell.edu.
Is the Customer Always Right? The Potential for Racial Bias in Customer Evaluations of Employee Performance

Abstract
With the encouragement of marketing scholars, many companies are tying employee incentives to customer ratings of satisfaction, service quality, or employee performance. One potential drawback to these practices is that customers' evaluations of employees—and, therefore, any associated rewards—may be biased by employee race. This possibility was examined in a restaurant setting. We found that customers rated the promptness and attentiveness of same race servers more favorably than different race servers, but there were no differences for assessments of server friendliness or appearance. The theoretical and practical implications of these findings are discussed.

Keywords
norm, racial bias, customer evaluations, employee performance, ratings

Disciplines
Applied Behavior Analysis | Food and Beverage Management | Race and Ethnicity

Comments
Required Publisher Statement
Is the Customer Always Right? The Potential for Racial Bias in Customer Evaluations of Employee Performance

Michael Lynn and Michael Sturman
Cornell University

With the encouragement of marketing scholars, many companies are tying employee incentives to customer ratings of satisfaction, service quality, or employee performance. One potential drawback to these practices is that customers’ evaluations of employees—and, therefore, any associated rewards—may be biased by employee race. This possibility was examined in a restaurant setting. We found that customers rated the promptness and attentiveness of same race servers more favorably than different race servers, but there were no differences for assessments of server friendliness or appearance. The theoretical and practical implications of these findings are discussed.

Satisfying customers is an important key to business success. Highly satisfied customers are more likely than less satisfied customers to purchase again (Szymanski & Henard, 2001), to spread positive word of mouth (Anderson, 1998), and to accept price increases (Anderson, 1996). As a result, increasing customer satisfaction leads to greater revenues (Gruca & Rego, 2005) and lower marketing costs (Reichheld & Sasser, 1990), which ultimately increase profits and shareholder value (Anderson, Fornell, & Mazvancheryl, 2004). As one way of increasing customer satisfaction, academics in both marketing and human resources have advocated tying management and employee incentives to customer ratings of satisfaction, service quality, or employee performance (Hauser, Simester, & Wernerfelt, 1994; Lambert, Sharma, & Levy, 1997; Ulrich & Brockbank, 2005). In fact, many firms are following this advice and are basing incentives for service personnel, salespersons, and managers on customer evaluations (Lambert et al., 1997; Ulrich & Brockbank, 2005).

Although the arguments for using incentives based on customer evaluations are compelling, this practice has one potential drawback that is rarely investigated or discussed: Customer evaluations may be biased. That is, customer evaluations may be affected by factors unrelated to the behavior of the employees whose incentives they affect (Wherry & Bartlett, 1982). At best, such biases could be random and could reduce the reliability of customer ratings as measures of employee performance. At worst, such biases could be systematic and could lead to unfair, or even unlawful, distribution of
incentives/rewards among employees. The purpose of this paper is to examine one such potential bias. Specifically, we investigate the effects of employee race and its interaction with customer race on consumers’ evaluations of customer-contact employees.

We examine the potential for racial effects on customer-driven evaluations of employee performance in the context of the service sector. While customer evaluation of employee performance is not limited to any one industry, it is particularly relevant to the service industry for two reasons. First, customer-contact workers play a larger role in the production of services than in the production of physical goods (Berry, 1980). In fact, services are often defined entirely in terms of the behavior of customer-contact employees, so the behavior of those employees is a particularly important determinant of customer satisfaction for services (Bitner, Booms, & Tetreault, 1990). Second, the intangible and customized nature of services makes the behavioral control of employees both more difficult and less appropriate for service workers than for other workers (Anderson & Oliver, 1987; Eisenhardt, 1985). Outcome control, in the form of incentives that are tied to customer ratings of satisfaction or service quality, is a more efficient and effective way of ensuring that customer-contact employees deliver good service. Thus, the service industry provides a valuable opportunity for studying race effects on customers’ evaluations of employee performance.

Literature Review

Social psychologists have argued that although overt racism has declined over time, racial prejudice persists and continues to manifest itself in subtle, unintentional ways that can be justified on nonracial grounds (Dovidio, Gaertner, Kawakami, & Hodson, 2002). According to this view, many people in the United States today consciously endorse nonracist, egalitarian views and values but, because they are exposed to negative beliefs and attitudes about racial groups in the larger society, come to unconsciously harbor those societal prejudices (Devine, 1989).

These people’s conscious endorsement of egalitarian views means that although they may express a preference for their own race, they disavow any negative thoughts or feelings about other racial groups, find racial prejudice and discrimination aversive, and are motivated to inhibit or suppress any prejudice and discriminatory tendencies they may see in themselves (Plant & Devine, 1998). As a result, these aversive racists display few signs of overt or “old-fashioned” prejudice and discrimination. However, their conscious executive cannot inhibit automatic, nonconscious discriminatory behavior, or even conscious discriminatory behavior that is subtle and appears to be based on nonracial factors. Thus, these aversive racists do racially discriminate, but mostly in subtle, unintentional ways.
Consistent with this theory of aversive racism, researchers have found that although people continue to express a preference for their own race (Nosek, Banaji, & Greenwald, 2002), negativity toward racial minorities has declined over time (Dovidio & Gaertner, 1986). Explicit and implicit measures of attitudes toward racial groups differ from one another with greater levels of implicit than explicit prejudice (Nosek et al., 2002). In addition, the effects of prejudiced implicit attitudes on behavior are stronger for spontaneous behaviors and those deliberate behaviors that can be justified on nonracial grounds (Dovidio & Gaertner, 2000; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Sargent, 2004).

The theory and research on aversive racism described previously suggest that consumers’ evaluations of employee performance will be biased in favor of same-race employees, particularly to the extent that those evaluations can be attributed to nonracial factors. In the evaluation of employees in a customer-service interaction, some dimensions of employee performance seem to be more situation-specific and less reflective of employees’ dispositions than are others. In the service industry, for example, server promptness seems more situational and less dispositional than does server appearance. Ratings of more situation-specific aspects of employee performance should be easier to attribute to nonpersonal (i.e., nonracial) factors than ratings of more dispositional aspects of employee performance, so the former should be less subject to self-censorship and more likely to reflect racial biases than the latter.

Support for the existence of racial discrimination by consumers is strong (Adelman, 2004; Juni, Brannon, & Roth, 1988), as is support for the existence of racial bias in peer and supervisory ratings of performance (Stauffer & Buckley, 2005). However, there has been little research examining racial bias in consumers’ ratings of employee performance. Furthermore, although there is directional support for the idea that racial bias in performance evaluations will be greater for situation-specific than for dispositional dimensions of performance, no statistically reliable effect has been reported (Kraiger & Ford, 1985; Sackett & DuBois, 1991). Thus, there is a need for research examining customer racial discrimination in ratings of employees on both situational and dispositional dimensions of performance.

These effects are explored in a reanalysis of a study that was first reported by Lynn et al. (2008). Lynn et al. distributed surveys to restaurant customers, asking them to provide information on their servers’ personal characteristics, their servers’ performance, the meal, the restaurant, and themselves. They reported nonsignificant correlations between server race, customer race, and an index of customers’ service ratings, but that was not the focus of their original analysis. Their focus was on racial discrimination and tipping—not racial biases in customer ratings of employee performance—so they did
not examine the predictors of service ratings in any depth. Furthermore, although tipping is supposed to reflect customers’ evaluations of service, the relationship between restaurant tips and customers’ service ratings is very weak (Lynn & McCall, 2000), so Lynn et al.’s analyses of race effects on tipping are not indicative of race effects on service evaluations.

Our reanalysis focuses on service evaluations, and in three respects goes beyond the simple correlations involving service ratings that Lynn et al. (2008) reported. First, we examine the interactive effects of customer and server race, rather than just their main effects on service ratings. Second, we use additional data to test the extent to which individuals perceive the dimensions of performance used in Lynn et al. as disposition versus situational. Third, we test the hypothesis that racial biases affect customers’ ratings of more situational dimensions of service ratings differently than more dispositional dimensions of service ratings.

Method

The data for this study come from two sources. First, we reanalyzed data collected in Lynn et al.’s (2008) study, where researchers approached dining parties that had just been seated for lunch at a restaurant located in the southern United States and asked them to complete a survey. That restaurant employed multiple White and Black servers and had a racially diverse customer base.

A total of 155 usable observations were obtained and retained for analysis: 89 from White customers waited on by a White server; 21 from Black customers waited on by a White server; 34 from White customers waited on by a Black server; and 11 from Black customers waited on by a Black server. Servers’ characteristics and performances were assessed by asking customers to rate how much they liked the server’s (a) appearance; (b) friendliness; (c) attentiveness; and (d) promptness on a 5-point scale ranging from 1 (dislike very much) to 5 (like very much).¹

Preliminary analyses of the separate service ratings indicated that racial biases affected server attentiveness and promptness more than server appearance and friendliness. To explain these findings, we hypothesize that server attentiveness and promptness are more situational dimensions of service, while server appearance and friendliness are more personal dimensions of service. To test this theoretical proposition, we asked 41 students and staff members at an Ivy League university to indicate how much they thought a restaurant server’s performance on each dimension is a result of situational factors outside the server’s control (e.g., how many other customers the server has, how complicated the order is, how well staffed the kitchen is) versus personal characteristics of the server (e.g., his or her

¹ For more details about the study methodology, including the treatment of missing values and other variables not used in this reanalysis, see Lynn et al. (2008).
ability, motivation, personality). Participants were asked to rate server promptness, server friendliness, server attentiveness, and server appearance, in that order, on a 9-point scale ranging from 1 (more situational) to 9 (more personal).

Results

A within-subjects ANOVA on the student survey data indicates that the different dimensions of service were perceived as differing in terms of how much they were affected by situational factors outside the server’s control versus personal characteristics of the server, $F(3, 120) = 55.81, p < .001$. More specifically, post hoc comparisons indicate that server promptness was perceived as significantly ($p < .001$) more situational ($M = 4.06, SD = 1.63$) than server attentiveness ($M = 5.85, SD = 2.08$). Similarly, server attentiveness was perceived as significantly ($p < .001$) more situational than either server appearance ($M = 7.68, SD = 1.78$) or server friendliness ($M = 7.78, SD = 1.47$). However, server appearance and server friendliness did not significantly differ from one another ($p > .50$). Thus, our hypothesis that server promptness and attentiveness would be perceived as more situational dimensions of service than would server appearance and friendliness was supported.

Using confirmatory factor analysis, we tested if Lynn et al.’s (2008) ratings of server appearance and friendliness (labeled personal dimensions of service) loaded on one factor, and the ratings of server attentiveness and promptness loaded on a second factor (labeled situational dimensions of service). We compared this factor structure to the null hypothesis that all four items loaded on a single construct. The confirmatory factor analyses show that the two-factor solution fit the data better than did the one-factor solution ($p < .0001$) and had good overall fit (root mean square error of approximation = 0.00; normed fit index = 1.00; comparative fit index = 1.00). Note also that the hypothesized two-factor solution was better fitting than were all other potential two-factor solutions. Accordingly, we averaged the two items underlying each factor and used those indexes in subsequent tests of racial biases in customers’ evaluations of employee performance. Coefficient alphas were .82 and .91, respectively, for the personal dimensions of service index and the situational dimensions of service index.

A 2 x 2 x 2 ANOVA on customer ratings of server performance using dimensions of service as a within-subjects factor and both server and customer race as between-subjects factors produced significant effects for service dimension, $F(1, 151) = 7.13, p < .009$; as well as for the three-way interaction between service dimension, server race, and customer race, $F(1, 151) = 4.15, p < .05$. The situational dimension of service was rated slightly but significantly lower than was the personal dimension of service (estimated marginal Ms = 4.34 vs. 4.50). Moreover, the two-way interaction
between server and customer race was significantly stronger for the situational dimension of service index, $F(1, 151) = 4.09, p < .05$; than for the personal dimension of service index, $F(1, 151) = 0.61, ns$.

<table>
<thead>
<tr>
<th></th>
<th>Situational service index</th>
<th>Dispositional service index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>White customer/White server ($n = 89$)</td>
<td>4.53</td>
<td>0.74</td>
</tr>
<tr>
<td>White customer/Black server ($n = 34$)</td>
<td>4.21</td>
<td>0.88</td>
</tr>
<tr>
<td>Black customer/White server ($n = 21$)</td>
<td>4.12</td>
<td>1.13</td>
</tr>
<tr>
<td>Black customer/Black server ($n = 11$)</td>
<td>4.50</td>
<td>0.63</td>
</tr>
</tbody>
</table>

When rating situational dimensions of service, restaurant customers showed a same-race bias such that White customers rated White servers’ performance better than Black servers’ performance ($M_s = 4.53$ vs. $4.21$), $t(121) = 2.64, p < .05$; and Black customers rated White servers’ performance lower than Black servers’ performance ($M_s = 4.12$ vs. $4.50$), $t(30) = 1.03, ns$. In contrast, when rating personal dimensions of service, restaurant customers showed no such racial bias (see Table 1).

Discussion

If we accept that racial bias should not influence the evaluation or reward of employees, then our results show that we cannot accept the adage “The customer is always right.” The results of this study clearly demonstrate the potential for race to affect customer evaluations of employee performance. While race effects were not observed on customer ratings of server personal characteristics (i.e., attractiveness, friendliness), a Customer Race ¥ Server Race interaction reflecting favoritism toward the customers’ own racial group was found on ratings of more situational dimensions of server performance (i.e., attentiveness, promptness). These findings have several implications for theory and practice.

Racial Biases in Customer-Driven Incentives

Our findings provide evidence that racial biases affect consumers’ evaluations of employee performance. The customer race by server race interaction effect on ratings of situational dimensions of server performance is consistent with widespread conscious preferences in the United States for
members of one’s own race over other races (Ashburn-Nardo, Knowles, & Monteith, 2003; Nosek et al., 2002). While previous research has shown that these preferences are reflected in consumers’ choices (Adelman, 2004; Juni et al., 1988), ours is the first study to extend these findings to customer evaluations of employee performance.

Our findings are particularly interesting because they contrast with the racial effects on tipping observed by Lynn et al. (2008), using the same data from the same restaurant customers about the same dining experiences. Lynn et al. found no race similarity effect. Rather, they found that both White and Black customers tipped Black servers less than they did White servers. The behavior of White customers seems consistent: They tip and rate White servers more highly than they do Black servers. However, Black customers rate Black servers more highly than they do White servers, but they tip the former less than the latter. Why? Perhaps Black customers feel less social pressure to tip Black servers than to tip White servers. Blacks and Whites appear to have different perceptions and behaviors regarding tipping (Lynn, 2006; Lynn & Thomas-Haysbert, 2003), so Black customers may think that their lower typical tip amounts will be more acceptable to Black servers. This possibility deserves further investigation in future research.

Practical Implications

On a practical level, our findings call into question the advisability of using customer evaluations as part of employee incentive systems. More explicitly, the results suggest that the use of customer-driven incentives may violate the Civil Rights Act of 1964, which prohibits employment discrimination on the basis of race, color, religion, sex, or national origin. In Griggs v. Duke Power Company (1971), the Supreme Court ruled that the Civil Rights Act of 1964 prohibits business policies and practices that have a disparate impact on protected classes of applicants and employees, even if those policies and practices appear at face value to be neutral and are not intended to discriminate (cf. Arvey & Faley, 1988; Twomey, 1998).

Customer performance ratings—when tied to individual employees and individual rewards—that favor same-race employees would have a disparate impact on the evaluations of any employee racial group that is underrepresented in the customer base. Thus, basing raises or promotions on customer evaluations of individual employee performance may qualify as such an apparently neutral business practice that has an unintended disparate impact on employees of different races. In order to avoid potential lawsuits, managers should check for such racial biases in customers’ appraisals of employee performance before they use those appraisals in compensation or promotion decisions, and
those customer appraisals should not be used if racial biases are evident. Alternatively, customer evaluations could be averaged to reflect group service levels, and thereby could be tied to group awards. This approach would result in equal pay across group members, but, as such, may not be as relevant to incentivizing the specific customer–employee interaction that managers want to evaluate and reward.

Despite the aforementioned warnings about the use of customers to evaluate employee performance, we want to be clear that we do not advocate the idea that their use be abandoned completely. Given the importance of customer satisfaction to business success (Anderson et al., 2004), the logic behind obtaining input from this important constituency is still compelling. We are suggesting that managers need to be aware of the possibility for racial bias in customer-driven incentives; they need to test for such bias in their own customer-driven incentives; and they need to take corrective action if such biases are found.

Limitations and Directions for Future Research

The results of this study, although useful as a demonstration that racial bias can affect customer-driven employee evaluations, are nonetheless limited in two respects. With respect to causal mechanisms, it is possible that the interaction in this study could reflect a tendency for servers to deliver better service to same-race customers, rather than a racial bias in customers’ evaluations. We cannot explicitly refute this alternative explanation with our data. However, we argue that this alternative is unlikely for several reasons.

First, servers are happier to receive large tips than small ones (Noll & Arnold, 2004) and tend to give better service to customers whom they perceive to be better tippers (Barkan & Israeli, 2004). Furthermore, servers of both races perceive Whites as better tippers than Blacks (McCall & Lynn, 2009; Noll & Arnold, 2004), and in this restaurant, White customers did tip substantially more than did Black customers, regardless of the race of the server (Lynn et al., 2008). Thus, both Black and White servers were more likely to favor White customers than same-race customers.

Second, if servers ignored the incentives provided by tipping, preferred waiting on same-race customers, and, therefore, gave those customers better service, then the customer race by server race effect should have been observed for evaluations of server friendliness, as well as server promptness and attentiveness. In fact, since both the encoding and decoding of friendliness is strongly affected by nonverbal behaviors that are difficult to control consciously (Argyle, 1988), any server preferences for same-race customers should have affected their friendliness even more than their attentiveness or
promptness, which servers can more consciously control. Contrary to this hypothesis, racial biases were significantly stronger for the latter variables than for the former. This differential pattern of results for ratings of server friendliness, as compared to ratings of server promptness and attentiveness, can be explained more readily in terms of consumer racial bias than in terms of discrimination in service delivery.

With respect to generalizability, the specific customer evaluations, service setting, and racial groups in our study limit the breadth of conclusions we can draw. Such limits are inherent in any single study and could not be avoided. Nevertheless, it is worthwhile considering them in more detail as they point to potential directions for future research.

First, only a small set of customer ratings was used in this study. Our ratings of server appearance, friendliness, attentiveness, and promptness were well suited to our theoretically derived (albeit post hoc) distinction between situational and personal dimensions of service, in keeping with our need for a short questionnaire that restaurant diners would actually complete, and consistent with the customer ratings of service in the tipping literature (cf. Conlin, Lynn, & O’Donahue, 2003). However, they do restrict our ability to generalize the findings to other commonly used measures of customer satisfaction or service quality (e.g., Parasuraman, Zeithaml, & Berry, 1985).

Second, the generalizability of our findings is limited by the fact that our data came from only a single restaurant. There are no a priori reasons for expecting this unit of a national restaurant chain to be unique, but it is possible that consumer racial biases are more pronounced in some service contexts and in some geographic regions than in others, so more research is clearly needed to replicate our findings in additional settings.

Finally, the generalizability of our results is limited by our focus on two races. The sample of customers and servers was not diverse enough to consider other racial groups in our analyses. This is not a critical failing, because two racial groups were sufficient to test our hypotheses; and the minority group we did study (i.e., Blacks) is a large and economically important one (Williams & Tharp, 2001). Nevertheless, racial diversity in this country is not limited to Blacks and Whites. In fact, Hispanics are now the largest and fastest growing ethnic minority group in the country (Nasser, 2003). Our analysis of consumer racial discrimination can and should be extended to a larger number of ethnic and racial groups, and should be tested in situations with more diverse racial representation.

In summary, this study found that restaurant patrons rated the performance of same-race servers higher than that of different-race servers, at least on more situational dimensions of service. This finding provides support for the basic tenets of aversive racism theory (Dovidio, Gaertner et al., 2002). In
addition, it suggests that Lynn et al.’s (2008) finding that Black customers tipped Black servers less than they did White servers is not a result of affective biases against their own race, as Lynn et al. suggested.

Furthermore, the results of this study suggest that the use of customer ratings as part of employee incentive systems may have an adverse impact that violates laws against racial discrimination in employment. Thus, managers should test for racial biases in their own customer-driven incentives and take corrective action if an adverse impact is found. Of course, these results reflect the behavior of patrons at only one restaurant, so they should be replicated in different settings. Finding settings with racially diverse employees and customers is challenging, but we hope that the serious nature of our findings encourages more researchers to undertake this task and more companies to cooperate in the investigation of these important research questions.

References


