2012

Implicit racial prejudice against African-Americans in balanced scorecard performance evaluations

David R. Upton

C Edward Arrington
University of Wollongong, edarr@uow.edu.au

Publication Details

Upton, D. R. & Arrington, C. Edward. (2012). Implicit racial prejudice against African-Americans in balanced scorecard performance evaluations. Critical Perspectives on Accounting, 23 (4-5), 281-297.

Implicit racial prejudice against African-Americans in balanced scorecard performance evaluations

Keywords
era2015

Publication Details
Upton, D. R. & Arrington, C. Edward. (2012). Implicit racial prejudice against African-Americans in balanced scorecard performance evaluations. Critical Perspectives on Accounting, 23 (4-5), 281-297.

Implicit Racial Prejudice Against African-Americans in Balanced Scorecard Performance Evaluations

ABSTRACT

A dominant theme in critical accounting theory has to do with the relation between the construction of human identities and accounting discourse and practices. Though this theme has strong antecedents in Marxist –inspired critique of ideology, research into accounting's role in the construction of identity has employed diverse approaches; among them, genealogical studies (e.g., Miller and O'Leary, 1987), deconstructive studies (e.g., Shearer and Arrington, 1993), psychoanalytic studies (Roberts, 1991 and 2009) and critical-rational studies (e.g., Power and Laughlin, 1996). This paper offers a different approach grounded in social-cognitive concerns with ways in which implicit attitudes about race influence our evaluation of others. To the extent that accounting is a ubiquitous medium for the evaluation of others, it is important to know if rationally irrelevant criteria do in fact influence such evaluations. This paper reports on the results of an empirical, lab-based study of balanced scorecard evaluations and bonus allocations where race is a treatment effect and where the well-established tenets of Implicit Association Testing (IAT) are used to reveal that there are, indeed, propensities to unwillingly let racial prejudice intervene into our accounting-based evaluations of others. That intervention influences identity in ways that are morally unacceptable, degrading to black workers, and loaded with potential for negative material consequences for workers (e.g., less compensation due to racially-determined and irrational performance evaluations).

Keywords: Implicit social cognition, balanced scorecard, performance evaluation, race bias

Implicit Racial Prejudice Against African-Americans in Balanced Scorecard Performance Evaluations

1. Introduction

> *To engage in a serious discussion of race in America, we must begin not with the problems of black people but with the flaws of American society – flaws rooted in historic inequalities and longstanding cultural stereotypes.*
>
> Cornel West, Race Matters, p. 6

This study examines whether or not white American MBA and MSA students demonstrate implicit racial prejudice and whether or not such prejudice is present in evaluating the performance of managers in a balanced scorecard setting. The importance of the study derives from the fact that accounting plays a significant role in the evaluation of persons for the quality of their work. Accounting evaluation, unlike many other forms of evaluation, has economic "muscle" in that accounting evaluations are used to allocate resources, to influence the nature of work, and to motivate certain kinds of behaviors that have powerful implications for the way in which day-to-day life is lived. Much research has been oriented toward several aspects of accounting's place in performance evaluation activities, though these aspects tend to be narrow, constrained to rational theories, and grounded in the desirability of capitalistic norms (e.g., treating workers as self-interested functionaries in need of control, incentives, and motivations). Such research has explored areas like the following: (1) the normative conditions that would make some forms of accounting more desirable than others (Merchant, 2006); (2) the salience of different types of accounting information to the performance evaluation process (e.g., Feltham and Xie, 1994; Baiman and Rajan, 1995; Moers, 2005); and, (3) how different types of contracts and financial reward systems mediate both performance evaluation and performance quality (e.g., Ittner and Larcker, 1998; Scott and Tiessen, 1999; Bonner and Sprinkle, 2002). In this study, we address performance evaluation in a balanced scorecard setting with experimental focus on race and the manner in which explicit and implicit attitudes about race influence performance evaluation decisions, and we do so in a manner more consistent with critical-theoretic and ethical concerns rather than functionalistic and capitalistic concerns.

In doing so, we open accounting research to what is perhaps the most formidable domain of contemporary social-cognitive psychological research, an area concerned with attitudes in general and implicit attitudes more particularly. Implicit prejudicial attitudes are "… automatically activated and occur without full awareness upon exposure to a Black person" (Dividio et al., 2009, p.166). The paper

examines the effect of both explicit and implicit attitudes about race on performance evaluation decisions and bonus allocations.

An *explicit* attitude is one about which its holder is fully aware. A racist who knows and embraces his or her racism holds an explicit racist attitude. Though more nuanced definitions are available, in general an *implicit* attitude is one about which the person who holds it is not consciously aware that he or she does in fact possess it (see Petty, Fazio and Briñol, 2009, pp. 3-9 for extensive discussion of the meaning of implicit attitudes). If attitudes – like other personality attributes – influence judgments and decisions, including decisions about other peoples' performance; and, if attitudes are implicit, then the outcome of accounting-based performance evaluations may be driven in part by unintended effects, effects that are both prejudicial at an ethical level and detrimental to organizational performance at a pragmatic level.[1]

Implicit racial biases and implicit gender biases are two outcomes that have been observed in a large number of studies (see Nosek et al., 2007b; Petty et al., 2009; Dovidio et al., 2009 for comprehensive reviews). This paper brings inquiry into implicit attitudes about race to bear upon performance evaluation in a balanced scorecard setting. The importance of understanding whether implicit and unintended racial attitudes influence performance evaluation goes without saying – legal, ethical, and political issues are all involved. The ethical implications for judging evaluators are very different across explicit racial bias and implicit racial bias. An explicit attitude of racial bias is one for which an evaluator is and should be accountable – he or she is simply a racist. Implicit racial bias is different: to the extent that ethical responsibility presupposes awareness, then implicit racial bias has less to do with ethical responsibility and more to do with the tacit consequences of the manner in which an evaluator's own identity has been constructed from within a social ontology where racial issues have been historically salient. The implicitly biased evaluator is simply unaware of that bias (see Messick and Bazerman, 1996; Banaji et al., 2003).[2]

As mentioned, we investigate the nature of racial bias as manifest in the performance evaluation and reward judgments of managers in an experiment based on the balanced scorecard (BSC hereafter). The BSC is recognized as an important mechanism to facilitate performance evaluations in organizations. Its advocates also recommend linking BSC-based evaluations to reward systems (Kaplan and Norton,

---

[1] Of course, we are not suggesting in any sense that prejudice would be "acceptable" if it enhanced organizational performance. We are simply suggesting that any prejudice that mediates performance evaluation does not contribute to organizational performance however it is construed.

[2] The "Bradley Effect" is related to but different from the concern with implicit racial bias. The Bradley effect is a well-cited phenomenon which, put simply, suggests that black candidates will perform better in opinion polls than they do in actual elections. Implicit bias probably plays a role in creating this effect, though to our knowledge we have no evidence on the extent to which the Bradley Effect is created by implicit rather than explicit racial prejudice (see Peery and Bodenhausen, 2009 for comprehensive discussion of the various factors influencing forms of racial prejudice).

1996). We measure racial attitudes and beliefs both implicitly and explicitly to investigate whether such biases influence these decisions.[3]

In our experiment, the racial identities (black or white) of two unit managers are manipulated.[4] Participants, acting in the role of senior managers, evaluate the unit managers' performances based on a comparison of common actual to target performance measures presented in a BSC format. All of our participants are white, and thus the validity of our results extends only to the racial attitudes of white Americans. We include only white American participants for two reasons. First, we did not have an adequate sample size to make inferences about non-white participants. Second, the study of whites seems important in its own right since about 80% of the U.S. population is white (http://quickfacts.census.gov/qfd/states/00000.html).

In addition to performance evaluations, participants also allocate a fixed bonus pool of $20,000 between the two unit managers. Key dependent measures are the difference in the quantitative scores for performance evaluation of the two managers and the bonus allocations across the two managers. Participants then complete tasks that measure explicit and implicit attitudes about racial identity. Participants were evening MBA and MSA students from a U.S. southeastern university. The average age of participants was 32, and they had on average about10 years' work experience. Thirty were male and 21 were female.

Results show that differences in performance evaluations arise depending on the race of managers. Specifically, when a white manager outperformed a black manager the difference in the two evaluations was greater than it was for either two white or two black managers. However, in sharp contrast, when a black manager performed better than a white manager a smaller evaluation difference was found compared to other pairings. We also find biases in performance among the four dimensions of the BSC, with greater bias appearing in the customer and the learning and growth dimensions. This seems consistent with some literature which suggests that implicit racial prejudice may be less forceful in evaluations based on "hard, quantitative, and analytic" concerns (like accounting numbers). Participants self-reported a favorable explicit attitude towards blacks, but results reveal implicit attitudes of preference for whites over blacks. Empirically, the implicit attitude measure predicted performance evaluations and bonus awards when the white manager outperformed the black manager but not when the black manager outperformed the white manager.

Implications of the study are twofold: 1) empirical evidence of implicit and biased attitudes toward blacks is observed in the study and influences both performance evaluations and bonus

[3] Greenwald et al. (2009a) recommend both implicit and explicit attitude measurement within studies because of their mutually incremental predictive ability, especially for socially sensitive domains like sexism and racism.
[4] We only study white American participants. Any results drawn from this study are not applicable to any other group.

allocations; and, 2) given that an implicit bias for whites over blacks may be present in actual performance evaluation settings, management must seek out methods to mitigate this bias in addition to any measures mitigating explicit bias (e.g., diversity, EEO, etc) to avoid both potentially costly discriminatory behavior and unjust moral degradation of the quality of life for some employees. As we discuss later, approaches to dealing with explicit bias (e.g., training sessions) may be ineffective for implicit bias. It seems that these sorts of prejudices are so very deeply grounded in the history of cultural practices that "methods" (like training) are too modest to do much good.

The remainder of the paper proceeds as follows. In the next section we discuss relevant literature on implicit bias and the BSC and propose our hypotheses. We then describe the experiment. Experimental results are then presented, and we conclude with a discussion of implications, limitations, and trajectories for future research.

2. Background and hypotheses development

2.1. Race bias in performance evaluations – field and lab findings

Discrimination based on race within organizations is still pervasive according to current research. For example, job interview callbacks (Bertrand and Mullainathan, 2004), new hires (Giulano et al., 2009), experiences within the accounting profession (Weisenfeld and Robinson-Backmon, 2007), and merit-based evaluations and reward systems (Castilla, 2008) are shown to favor whites over blacks. Because of the commonness of such prejudice, blacks will sometimes even hide their race in order to circumvent perceived racial injustices (Luo, 2009).

In relation to studies of race bias in job performance a formidable record of findings has emerged over an extended period of time. Meta-analyses report the basic finding that performance evaluations favor white over black managers (Kraiger and Ford, 1985; Pulakos et al., 1989; Waldman and Avolio, 1991; Roth et al., 2003; McKay and McDaniel, 2006). McKay and McDaniel (2006) identify the most influential moderators of the black-white mean difference in work performance as criterion type (e.g., job knowledge, work samples) and cognitive loading of criteria (compared with personality loading criteria). Performance measures requiring or related to cognitive ability tended to have larger effect sizes than contextual, personality-loaded criteria. Objective versus subjective performance evaluations showed equivocal results.

Intuitively, it seems that performance evaluation and reward systems based on merit may counteract race bias and therefore mitigate the negative consequences of such inequitable evaluation practices. However, in a field study of an organization operating a merit-based performance evaluation

and reward system, Castilla (2008) identifies opportunity structures that harbor discrimination. Castilla (2008) investigates a form of evaluative discrimination (Peterson and Saporta, 2004) and suggests that discrimination can be present as either 1) performance evaluations bias (bias in performance ratings); or, 2) performance reward bias (bias in the link between performance evaluations and career outcomes, such as salary increases and promotions). Castilla (2008) finds performance reward bias after controlling for merit and concludes that even under a merit-based system biases can emerge when accountability and transparency are limited.

2.2. Implicit Social Cognition and Its Measurement

Implicit social cognition theory examines implicit attitudes and stereotypes that are predictive of behavior in settings where explicit attitudes lack predictive ability (Greenwald and Banaji, 1995). "The signature of implicit cognition is that traces of past experience affect some performance, even though the influential earlier experience is not remembered in the usual sense – that is, it is unavailable to self-report or introspection" (Greenwald and Banaji, 1995, pp.4-5). This awareness is at the core of our concern for the role of racial prejudice in the social ontology of selfhood. The theory argues that the mind operates at two levels: a conscious level and an unconscious level (Fazio, 1990; Greenwald and Banaji, 1995; Wilson et al., 2000). Implicit attitudes reflect automatically activated evaluative information whereas explicit attitudes reflect evaluations produced by controlled processes (Greenwald and Banaji, 1995; Wilson et al., 2000; Nosek, 2005). The sense in which unconscious processes are salient is certainly not "news" to many who are well aware of the dynamic and rich work in psychoanalytic, Marxist, Nietzschean, and other approaches deeply relevant to both psychology and social ontology.

Experimental research in the area of attitude-behavior relationships has traditionally relied on self-report methods such as Thurstone, Likert, and semantic differential scales for the measurement of attitudes and beliefs (Banaji and Heiphetz, 2010, p.355). While explicit measurement methods are important since they provide a rationale for explicit justification of decisions and actions, they have limitations where for self-presentation concerns or for other reasons respondents do not wish to reveal their preferences or beliefs. Early research into attitudes did not have the methodologies which would make testing of implicit attitudes robust; but, as Petty et al. (2009) remark, testing of what is broadly termed "automatic" (that is, precognitive) associations is ubiquitous now:

> . . . today it is possible to assess automatic evaluation reactions. It is difficult to know exactly what was the first shot fired in establishing the current wave of research on automatic measures of attitudes, but it likely had to do with the growing acceptance of the idea that attitudes could be characterized as object-evaluation associations in memory that could vary in their accessibility (see Fazio, 1995). Furthermore, the voluminous research on associative priming in cognitive psychology (e.g., *doctor-nurse*; Meyer and Schvaneveldt, 1971) inspired efforts in social

> psychology to examine the automatic associations people had to social objects rather than just physical ones. (p. 5)

For implicitly held attitudes and stereotypes that are not easily accessible via conscious processes, self-report methods are simply unable to capture implicit social cognition (Greenwald, et al., 2009a). Although the correlation between self-report measures and behaviors is well established (Kraus, 1995), the relationship is not always strong, and more can be learned by examining the role of implicit social cognition. Recent developments in the measurement of implicit social cognition have been the catalyst for renewed interest in studying the effects of attitudes and stereotypes on behavior. The implicit association test (IAT) is one method to measure implicit social cognition, and it has been used extensively in research studies.[5] The IAT is an indirect method of measuring attitudes or stereotypes (Greenwald et al., 1998). It has been the methodological basis for hundreds, if not thousands, of studies. Numerous examples of the ways in which it has been experimentally deployed are available at the following website: *www.implicit.harvard.edu.* We employ the IAT here.

In the race IAT, participants sort exemplars of two target concepts (European American (EA) and African American (AA)) and two attributes ('Good' and 'Bad') with two keys on a computer (in our experiment the two keys were the 'd' key and the 'k' key). The operational logic is that faster associations (for example, between EA+Good compared with AA+Good) are indicative of a stronger preference or belief for the first target concept. Table 1 specifies the sequence of blocks and number of trials used in the IAT. The test blocks are 3, 4, 6, and 7. The task in these blocks is to sort the four categories using the two keys. In blocks three and four (the compatible pairing) one key represents a pairing of EA + Good, and the other key represents a pairing of AA + Bad. In blocks six and seven (noncompatible pairing) the pairings are switched to AA + Good and EA + Bad. A participant who responds more quickly in the compatible trials compared to the noncompatible trials is said to have a preference for whites (i.e., a faster response implies a stronger association of EA + Good compared to EA + Bad). The IAT produces a D score, which is a latency response measure. It is computed as the difference in average response latency between the IAT's two combined tasks (e.g., EA + Good, EA + Bad), divided by an "inclusive" standard deviation of subject response latencies in the two combined tasks (see Greenwald et al., 2003 for more instrumental explanation).

[Insert Table 1 about here]

[5] The IAT and priming (Fazio et al., 1995) are the most commonly used methods to measure implicit social cognitions (Lane et al. 2007). Other techniques include the Go/No-Go Association Task (Nosek and Banaji, 2001), the Evaluative Movement Assessment (Brendl et al., 2005), the extrinsic affective Simon task (De Houwer, 2003), and the affect misattribution procedure (Payne et al., 2005).

Implicit social cognition assumes both scientific and practical importance when it impacts behaviors of interest differently from explicitly measured attitudes or beliefs (Greenwald and Krieger, 2006). The first study to show that implicitly measured bias is related to intergroup discrimination was reported by McConnell and Liebold (2001), though interest in implicit attitudes extends back some sixty years (Hovland, Janis and Kelley, 1953). In McConnell and Liebold (2001), the behavior of white students was evaluated as they interacted with a white and a black interviewer. Measures included ratings on friendliness, curtness, body language, social distance, and speech patterns. Participants also completed implicit and explicit measures of race attitude. Stronger implicit race bias was related to more discriminatory social interactions. The implicit and explicit measures were also positively related. Implicit race bias has been correlated with behaviors in many other settings (see Dovidio et al., 2009). For example, Rudman and Lee (2002) found a positive relation between implicitly measured racial stereotyping and judgments of hostility and sexism, implicit race attitude was predictive of voting in the 2008 presidential election (Greenwald et al., 2009b), and in a web-based experiment prowhite implicit bias increased the likelihood of treating white patients and not treating black patients with thrombolysis (Green et al., 2007 p.1231).

More generally, the predictive ability of implicit attitudes has been evidenced in a wide range of settings in addition to race. In a meta-analysis Greenwald et al. (2009a) report positive correlations between implicit attitudes and criterion measures covering domains from race, to consumer and political preferences, to personality traits. In most domains self-report measures were more predictive of behavior than implicit measures. However, in the race domain the correlation between implicit attitudes and behavior was stronger than between self-report measures and behavior. This is immediately suggestive of the social undesirability of racial attitudes inasmuch as subjects who do possess explicit attitudes indicative of racial prejudice are reluctant to report those attitudes. While early twentieth-century racism tended to be overt (see Dovidio, 2001), the situation today is different and more complex:

> . . . contemporary racial prejudice is more complex and expressed more subtly than traditional prejudice. . . . In contrast to "old-fashioned" racism, which is blatant, aversive racism represents a subtle, often unintentional form of bias that characterizes many White Americans who possess strong egalitarian values and who believe that they are nonprejudiced. Because of the central role that racial politics has played in the history of the United States, this research has focused on the attitudes of Whites toward Blacks. Nevertheless, we note that many findings and principles we discuss extend to biases toward other groups, such as Latinos (Dovidio, Gaertner, Anastasio, and Sanitioso, 1992) and homosexuals (Hebl, Foster, Mannix and Dovidio, 2002). (Dovidio et al., 2009, p. 167)

The concept of "aversive racism" above captures a mode of racial prejudice that – like most postcolonial phenomena – is less visible than earlier forms of prejudice and political oppression. This aversive racism works subtly, invisibly, and in a manner whereby those who possess it are unaware of

their own racist propensities. Dovidio (2001) explains aversive racism in the context of a "third wave" of research into racial prejudice:

> The third wave of research on prejudice, which began in the mid-1990s and characterizes much of the current research, emphasizes the multidimensional aspect of prejudice and takes advantage of new technologies to study processes that were earlier hypothesized but not directly measurable. For example, whereas aversive racism, modern racism, and symbolic racism approaches to contemporary prejudice assumed the existence of widespread unconscious negative feelings and beliefs by Whites toward Blacks, new conceptual perspectives (e.g., Greenwald and Banaji, 1995) and technologies (e.g., response latency procedures; Dovidio and Fazio, 1992; Greenwald, McGhee, and Schwartz, 1998) were developed to understand and measure *implicit* (i.e., automatic and unconscious) attitudes and beliefs. These new technologies permit the assessment of individual differences in implicit, as well as explicit, racial attitudes and may thus help distinguish traditional racists from aversive or modern racists from truly nonprejudiced Whites (pp. 832-33).

Research shows that implicit attitudes may develop from early experiences, affective experiences, cultural biases, and cognitive consistency principles (Rudman, 2004). For example, Baron and Banaji (2006) found that white Americans' implicit and explicit attitudes towards blacks diverged from childhood to adulthood. That is, six-year olds held similar implicit and explicit attitudes; however, through age 10 and into adulthood, the implicit bias remained but the self-report attitude softened. In a major web-based study Nosek et al. (2007b) report evidence of the pervasiveness of implicit bias across a wide range of criterion measures. Their data on race bias shows a preference for whites over blacks for 68% of participants.

Chugh (2004) discusses the implications of implicit social cognition for managers. Chugh (2004) argues that the distraction, time-pressure, and ambiguity inherent in managerial roles establish conditions under which implicit biases can influence managerial behavior. In a broad sense, stress, overwork, and time pressure make rational cognitive processes less available, and this can open a space for less rational influences on decisions. Managers frequently use multiple performance measures systems, such as the BSC, to evaluate performance.[6] Potentially this is a setting where implicit attitudes or beliefs could influence managerial judgments even when managers explicitly disavow such influences.

2.3. Balanced Scorecard

The Balanced Scorecard (BSC) is a rather recent development in managerial accounting (Ittner and Larcker, 2001). Kaplan and Norton (1992) introduced the BSC as a performance measurement and reward system that helps link operational performance measures to the implementation and monitoring of strategy (1996, 2000). The BSC typically consists of four sets of measures: financial, customer, internal

[6] The BSC is extensively used by companies worldwide (Rigby and Bilodeau, 2009).

processes, and learning and growth. Though the BSC is a compelling innovation because it incorporates strategy, process, and managers to provide an integrated system of planning and control (Atkinson et al., 1997, pp.93-94), there is really very little credible academic evidence that it delivers enhanced performance, even on its own value-terms (see Nørreklit, 2000).The BSC does however provide a breadth and scope of performance evaluation that goes beyond financial metrics. As a caveat, and as is obvious, the BSC is grounded in the rather narrow, functionalist goals of organizations oriented toward the efficacy of work and thus has little regard for other "outcomes" of work (e.g., the pathologies of subjective identity and moral experience which "work" and "evaluations" create).

Researchers have examined various explicitly cognitive biases found in the use of the BSC. Lipe and Salterio (2000) identified a common measures bias in the BSC where superiors ignore unique performance measures in favor of measures common to subordinates being evaluated. Subsequent studies have examined approaches to mitigate the common measures bias such as the use of strategically linked performance measures and strategy maps (Banker et al., 2004; Humphreys and Trotman, 2010), knowledge and training (Dilla and Steinbart, 2005), assurance and process accountability (Libby et al., 2004), and disaggregation of the assessment process (Roberts et al., 2004). Other biases examined when the BSC is used for performance evaluation include a 'selective attention to strategy effectiveness' bias (Wong-on-Wing et al., 2007), likeability bias (Kaplan et al., 2008), bias from BSC information organization (Lipe and Salterio, 2002; Cardinaels and van Veen-Dirks, in press), and evaluator ambiguity intolerance (Liedtka et al., 2008). Cognitive biases have also been considered in the use of the BSC for strategy development and evaluation. Tayler (2010) examines the influence of motivated reasoning on projects evaluated under the BSC.

The task of evaluating performance using the BSC has been characterized as both complex and ambiguous (Lipe and Salterio, 2000, 2002, p.532; Merchant, 2006). Nonfinancial performance measures are potentially more noisy, subjective, and ambiguous than traditional financial measures (Ittner and Larcker, 2003). Determining appropriate weights for nonfinancial measures is challenging because of validity and reliability concerns (Luft, 2009), and field research reports instances where weights on nonfinancial measures have been significantly reduced in part because of reliability concerns (Malina and Selto, 2001; Ittner et al., 2003). Often, evaluators will assign their own weights to these performance measures in arriving at overall evaluations (Ittner et al., 2003). Additionally, the cognitive effort required to assimilate multiple measures from the four BSC perspectives adds to the challenge of effectively employing the BSC. Lipe and Salterio (2002, p.533) suggest that evaluators will use a 'divide and conquer' approach, evaluating measures within the four perspectives and then determining an overall evaluation, as a strategy for managing the cognitive task. Moers (2005) discusses additional concerns with the use of multiple performance measures systems, such as the BSC, for performance evaluations. His

empirical study indicates that 1) performance measure diversity leads to more lenient performance ratings and less differentiation amongst employees; and, 2) more subjectivity leads to compressed and higher ratings.

Since ambiguity has been identified as a moderator in the relationship between implicit attitudes and behaviors (Chugh, 2004), the ambiguity produced through the measurement and cognitive challenges of the BSC increases the possibilities that the BSC enables implicit biases to affect performance evaluations. Racial prejudice – implicit or explicit – certainly influences a ubiquitous domain of moral, social, and political dimensions of life. In this study, such influences are directed squarely at the identity of employees: the overriding purpose of performance evaluation is to *evaluate* a person and thus to participate in the construction of that person's economic and organizational identity. In addition, the financial, customer, internal business processes, and learning and growth dimensions of the BSC include measures with varying degrees of validity and reliability. There is potential for implicit biases to surface to different extents across the four dimensions as well as in overall performance evaluation and reward decisions, largely due to differences in the decision-space associated with evaluator subjectivity across the four dimensions. One's subjectivity can, for example, embed itself more easily in a discourse of "learning" than it can in an algorithm of "financial performance." Further, as discussed in McKay and McDaniel (2006), the nature of measures within the four BSC perspectives may lend themselves more to cognitive versus personality loaded criteria. Recall that cognitive criteria were found to be more influential as moderators of black-white mean differences in performance.

2.4. Hypotheses

Our hypotheses are predicated on the selection of white American senior managers. The first two hypotheses address the possibility that discrimination based on race will be evident in performance evaluations and bonus allocations, indicative of either implicit or explicit prejudicial attitudes. Evaluators, or senior managers, may (unconsciously) bias performance evaluations when subjectively evaluating subordinates based on the BSC. Considering the context of this study – white American managers and whether or not they demonstrate racial bias -- biased evaluations would favor a white manager and would disadvantage a black manager. A similar pattern of biases is expected when a fixed bonus is awarded to managers based on their BSC performances. Thus our first hypotheses (assuming a white superior manager) are stated as follows:

> H1: Performance evaluation difference is highest when a white manager outperforms a black manager and lowest when a black manager outperforms a white manager.

H2: Bonus allocation to manager A is highest when manager A is a white manager who outperforms a black manager and lowest when manager A is a black manager who outperforms a white manager.

Implicit social cognition theory suggests that implicitly measured attitudes will reflect discriminatory performance evaluations and bonus allocations. Self-report measures of attitudes that are influenced by introspection and self-presentational concerns will not reflect discriminatory performance evaluations and bonus allocations. These relations should apply when white and black managers are compared but not when managers of the same race are compared. Therefore the third and fourth hypotheses are stated as follows:

H3: There is a positive correlation between performance evaluation and implicitly measured attitudes when a white manager outperforms a black manager and a negative correlation when a black manager outperforms a white manager.

H4: There is a positive correlation between bonus allocations and implicitly measured attitudes when a white manager outperforms a black manager and a negative correlation when a black manager outperforms a white manager.

3. Method

3.1. Overview and task

Hypotheses are examined through a 4x1 between-subjects experimental design. Four combinations of white (w) and black (b) managers form four RACE conditions as follows: two white managers (ww), two black managers (bb), one white and one black manager (wb), and one black and one white manager (bw). Participants, who were white American evening MBA and MSA students, assume the role of senior managers in the case and evaluate the performance of the two managers based on the BSC. Upon completion of the case task we measure both the explicit and the implicit attitudes of participants toward race.

The case is based on the BSC case used in Lipe and Salterio (2000).[7] In this case participants read how a company introduced the BSC. Participants are asked to assume the role of a senior manager and evaluate the performance of two unit managers. We modified the Lipe and Salterio (2000) case so that all the BSC measures were common and all targets were identical for the two business units (A and B), both of which are described as operating in retail markets (Kaplan, et al., 2008). Unit A manager outperformed unit B manager on all measures, and the actual performance of both managers exceeded all the targets. Further, the relative performance of unit A manager was consistently above that of unit B manager for all BSC perspectives (i.e., on average within each BSC perspective performance above target of unit A manager exceeded performance above target of unit B manager by the same ratio). Participants rated the performance of the two unit managers on each perspective of the BSC and on overall performance on a 100-point scale (with end-points of 'reassign' and 'excellent'). Participants also allocated a $20,000 bonus between the unit managers. Table 2 presents balanced scorecards for the two managers.

[Insert Table 2 about here]

3.2. Participants

Participants were white American graduate students in either evening MBA or evening MS in accounting classes enrolled at a U.S. southeastern university.[8,9] We do not include non-U.S. born participants or non-white U.S.-born participants in the analyses that follow since the biases hypothesized relate to white U.S. managers; therefore we use only white U.S.-born participants.[10] Further, as Dovidio et al. (2009) note, race is integral to US history in a way that probably is somewhat unique and certainly culturally powerful.[11] Participants on average had more than 10 years work experience and an average age of about 32. There were 30 males and 21 females in the study.

---

[7] Our thanks to the authors for supplying us with the case material.

[8] MBA and MSA participants had the same level of agreement concerning various aspects of the case, such as, how performance measures were categorized, the fact that the same performance measures were used for both units, the ease of completing the case, and case realism (all P-values>0.15). Therefore we combined results of the two groups in the analysis that follows. Participants earned extra course credit for participating in the experiment.

[9] Evening MBA and MSA participants surrogate for professional managers. These participants had on average 10 years work experience and prior studies rely on similar student participants to proxy for managers (see e.g., Lipe and Salterio 2000, 2002; Banker et al. 2004; Kaplan et al, 2008; Tayler 2010).

[10] All students in the MBA and MSA classes recruited for the experiment were invited to participate in the experiment. No indication regarding the nature of the race manipulation was given in the recruitment text. Participants not included in the white U.S. born category completed the experiment, but their data are not included in the analyses.

[11] Some may view a claim about the "uniqueness" of race and racism in an American context as problematic, jingoistic and wrong. We disagree. From the colonial period on, legalized racism, slavery, a Civil War, Post-war Reconstruction, the Civil Rights movement, culture and the scope of African-American immigration (forced or voluntary) make the shape and character of U.S. racism – if not unique – then at least culturally specific. Without further elaboration, we refer the reader to West (2001).

3.3. Procedure

The experiment has two phases. In phase one, participants completed three tasks in sequence: 1) the BSC case, 2) a post-experiment questionnaire, and 3) two self-report race attitude instruments. In phase two participants completed a computer-based Race Implicit Association Test (Race IAT), described to participants as a sorting task.

Consistent with many other studies, operationally we manipulated race by assigning recognizably black or white names to the unit managers (Bertrand and Mullainathan, 2004). These names were referred to repeatedly throughout the case materials. We used the names Greg Baker and Brad Walsh for the white managers and Darnell Washington and Tyrone Robinson for the black managers. Bertrand and Mullainathan (2004) establish the construct validity of these names as indicative of racial identity and discounted the possibility that these names indicate class.

3.4. Dependent Measures

3.4.1 Performance Evaluation

We calculate difference scores for the overall performance evaluation (DIFFOall) and each of the four BSC perspective performance evaluations (DIFF1 – DIFF4) by subtracting the evaluation of unit B manager from the evaluation of unit A manager. Positive scores indicate a higher evaluation for unit A manager than for unit B manager. We use the bonus allocated to unit A manager (BonusA) as the dependent measure for performance reward. Since participants allocated a fixed sum ($20,000) between the two managers, the bonus allocated to unit B manager is strictly determined by the bonus allocated to unit A manager.

3.4.2. Implicit Race Attitude Measure – Race IAT

Cropped faces of three female and three male European Americans and African Americans are used in the study (images source: webpage http://projectimplicit.net/nosek/). For exemplars for the attribute category "Good" we used love, peace, joy, wonderful, laughter, pleasure, glorious, and happy; for the attribute category "Bad" we used terrible, horrible, nasty, failure, agony, evil, awful, and hurt. The order of these pairings is counter-balanced in the experiment. We adapted the Race IAT provided by Greenwald (webpage: http://faculty.washington.edu/agg/iat_materials.htm) and ran the Race IAT using software by Millisecond. The IAT results in a D score that has a possible range of -2 to 2 with a score of zero indicating indifference between whites and blacks. In our experiment positive Race IAT scores

reflect a stronger preference for whites over blacks. We use the scoring algorithm proposed by Greenwald et al. (2003).

Evidence on the psychometric properties of the IAT has accumulated quickly since its introduction in 1998 (see, for example, Lane et al., 2007; Nosek et al., 2007a; Greenwald et al., 2009a). Internal consistency and test-retest reliability are considered satisfactory. Evidence of convergence validity between the IAT and other implicit measures is mixed, limited by the reliability of some implicit measures. Relations between the IAT and self-report measures vary as a meta-analysis showed an average r of .24 (Hofmann et al., 2005). Evidence of convergent and discriminant validity between IAT and self-report measures across attitude domains is judged to be strong. Predictive validity varies depending on context with the IAT showing greater predictive ability over self-report measures when studies involve discrimination toward a social group (Greenwald et al., 2009a).

3.4.3. Explicit Race Attitude Measures

Two traditional, commonly used self-report race measures were administered in the experiment (Devine et al., 2002; Payne et al., 2008). The first is the 'Attitude Towards Blacks' scale (Brigham, 1993) that consists of 20 items. Sample items include "African American and European American people are inherently equal" and "It is likely that African Americans will bring violence to neighborhoods when they move in." Participants respond to statements on a 7-point scale ranging from 1 'Strongly Agree' to 7 'Strongly Disagree'. After appropriate reverse scoring the items showed good reliability (alpha = 0.91); thus we created an explicit measure (ATB) by taking the mean score across the 20 items for each participant, with higher scores indicating a more negative attitude towards blacks.

The second self-report measure is a 'Feeling Thermometer' (introduced by Campbell, 1971 and used extensively, see for example, Amodio and Devine, 2006; Green et al. 2007; Payne et al. 2008). Participants report their feelings of coldness or warmth towards various racial groups on a scale from very cold (0) to very warm (10). We subtract the score for African Americans from the score for European Americans to arrive at a difference score (FT) of participants' preferences for whites compared to blacks. A positive score indicates a preference for whites over blacks.

The ATB scale is considered a cognitive, absolute measure in that it elicits participants' responses to statements concerning blacks only. In contrast, the feeling thermometer asks participants for their feelings towards various races and is thus an affective rather than cognitive scale. The FT difference score creates a relative scale contrasting blacks and whites. Conceptual correspondence is a feature of the FT as it parallels the affective, relative nature of the IAT. Evidence shows greater correlation between implicit and explicit attitude measures that have greater conceptual correspondence (Hofmann et al., 2005; Payne et al., 2008).

4. Results

4.1. Manipulation checks

The effectiveness of the race manipulation in the experiment depends upon the ability of participants to infer the race of the unit managers from their recognizably race-related names. Our manipulation check to secure the integrity of the manipulation asked participants to indicate the race (and gender and age) of five individuals based solely on their names. In the manipulation check we used a black name (Jermaine Williams) and a white name (Neil Murphy) drawn from the same list from which the case names were acquired. Bertrand and Mullainathan (2004) identified these names as being highly recognizable according to race. We did not use these names in the case materials. If participants correctly classified these names, then we assumed they also understood the race manipulation.[12] No U.S.-born white participants incorrectly classified the black and white names according to race (some participants responded with 'cannot tell').[13]

Participants in all conditions rated unit A manager's performance above that of unit B manager's performance, consistent with case BSC's that indicated unit manager A outperformed unit manager B on all measures.[14]

Experimental results are analyzed in two parts. First, we investigate if performance evaluations (H1) and bonus allocations (H2) differ based on race across the four RACE conditions (i.e., combinations of white (w) and black (b) managers: ww, wb, bw, bb) using one-way ANOVAs and planned contrasts. Second, using correlation analyses we examine if implicit and/or explicit race attitudes predict performance evaluations within the mixed race conditions (i.e., wb and bw conditions only).

4.2. Part I –Performance Evaluation Race Bias

Table 3 Panel A presents the means for performance evaluation differences (DIFFOall – performance evaluation for unit manager A minus performance evaluation for unit B manager) by RACE.

---

[12] We did not directly ask for the race of the unit managers because participants have been reticent to acknowledge the assumption of race based on names (Rudman and Lee, 2002).

[13] Specifically, 8 of 51 participants reported 'cannot tell' for the white name and 4 of 51 participants reported 'cannot tell' for the black name. Two white American participants were removed from the analyses because of missing performance evaluation and bonus allocation data, and one participant was removed because of an error rate of at least 25% in all the combined scoring blocks (see Nosek and Hansen, 2008).

[14] Participants agreed that the performance measures were usefully categorized ($p < 0.01$) and that it was appropriate to employ the same performance measures for both units ($p<0.01$), compared to the scale midpoints. There were no differences amongst participants across the four RACE conditions regarding the ease of understanding the case, the difficulty of the case, or case realism (all P-values > 0.210).

When managers of both unit A and unit B are white (ww) the difference in performance evaluations (DIFFOall) is 12.36. When managers of both unit A and unit B are black (bb) DIFFOall is 11.17. When unit A manager is white and unit B manager is black (wb) DIFFOall is 15.29. When unit A manager is black and unit B manager is white (bw) DIFFOall is 8.55.

[Insert Table 3 about here]

H1 states that performance evaluation differences (DIFFOall) in the wb cell will be greater than in the ww and bb cells, which will be greater than performance evaluation differences in the bw cell. Table 3 Panel B presents a one-way ANOVA with DIFFOall as the dependent variable and RACE as the independent variable. Results show a significant difference in means across conditions, and planned contrasts indicate significant differences as hypothesized.[15] Thus H1 is supported.

The H1 result suggests some race bias in the performance evaluations. Recall that in all race conditions the BSC performance measures and targets are identical for the unit A manager and the unit B manager and that the unit A manager outperforms the unit B manager by the same margin. Only the race of the managers differs across conditions. When a white manager outperforms a black manager participants assigned a large difference in evaluations between the managers, but when a black manager outperforms a white manager participants assigned a small difference in evaluations. When race was not a factor there was no difference in evaluations.

Table 4 Panel A presents the means for bonus allocation to unit A manager (BonusA). Means for the bonus allocated to unit A manager are as follows: ww (M=12,357), bb (M=12,183), wb (M=11,786), and bw (M=12,455).

[Insert Table 4 about here]

H2 states that the bonus allocation in wb is higher than in ww and bb, which is higher than the bonus allocation in bw. Table 4 Panel B presents a one-way ANOVA with BonusA as the dependent variable and RACE as the independent variable. Results indicate that the bonus allocations amongst race conditions did not differ. Further, planned contrasts indicate no bonus allocation differences between race conditions. [16] Therefore H2 is not supported.

---

[15] Specifically, a comparison of wb with ww and bb is significant (contrast = 3.52, p=0.035, one-tailed), a comparison of ww and bb with bw is marginally significant (contrast = 3.22, p= 0.062, one-tailed) and a comparison of wb with bw is significant (contrast = 6.74, p=0.003, one-tailed).

[16] Specifically, a comparison of wb with ww and bb is not significant (contrast = -485, p=0.904, one-tailed); a comparison of ww and bb with bw is not significant (contrast = -184, p= 0.677, one-tailed); and, a comparison of wb

Although performance evaluations based on the BSC result in significant differences among race conditions in favor of the white manager over the black manager, these differences do not carry over to bonus allocations. Interestingly, the bonus allocated to the white manager in wb cell is numerically the lowest of the four cells. However, while there is an absolute difference in dollar allocations across the experimental cells (ww, wb, bw, bb), that absolute difference is not statistically significant. We have little intuition into why this counterintuitive result prevailed other than the fact that a shift toward a language of dollars and away from a language of personal performance seems to have made a salient difference to the participants. Castilla (2008) revealed a performance–reward bias -- higher salary increases for whites over blacks, but in a decoupled reward decision – which may have some relation to this finding.

4.3. Bias in performance evaluations within perspectives of the BSC

A defining feature of the BSC is the classification of performance measures into four related perspectives. We asked participants to evaluate unit managers' performance on each perspective before evaluating the overall performance of the unit managers. Roberts et al. (2004) disaggregated the BSC by asking for evaluations on *individual* performance measures to address the common measures bias. Our approach reduces the cognitive effort required to evaluate all measures simultaneously (Lipe and Salterio, 2002) and is consistent with reported practice (Ittner et al., 2003). The following analyses investigate if race bias emerges differently under the four perspectives. Table 5 presents mean differences in performance evaluations for each perspective.

[Insert Table 5 about here]

We conducted four one-way ANOVAs with performance evaluation differences for each of the four perspectives (DIFF1…DIFF4) as the dependent variable and RACE as the independent variable. For the Customer-Related and Learning and Growth BSC perspectives marginally significant differences in means were indicated. Planned contrasts show that for both the Customer-Related and Learning and Growth perspectives performance evaluation differences (i.e., DIFF 2 and DIFF4) are higher for wb compared to ww and bb, and also compared to bw. . However the performance evaluation differences for ww and bb are not higher than for bw.

---

with bw is not significant (contrast = -669, p=0.930, one-tailed). We conducted planned comparison tests even though the RACE effect was not statistically significant because the hypotheses address specific relations (Keppel, 1991, p.112).

Performance evaluation biases were manifest in the Customer-Related and Learning and Growth perspectives and reflected in the overall performance evaluation differences. Arguably these perspectives consist of more ambiguous, or softer, performance measures than either the Financial or Internal Business Process perspectives. As stated earlier, it may be that with this greater subjectivity there is greater opportunity for evaluation biases to emerge. This result is compatible with Johnson et al. (2008) who found that gender bias was more likely to emerge from informal rather than formal measures of performance. This sort of finding has some very interesting implications for accounting research. It may well be that the formalism of accounting somehow dilutes the effect of prejudicial forces.

4.4. Part II relation to implicit and explicit measures of bias

The next set of analyses investigates the extent to which either implicitly or explicitly measured attitudes predict bias in BSC performance evaluations when white and black managers are evaluated. Descriptive statistics for the implicit and explicit measures are presented in Table 6, and correlations among the measures are reported in Table 7. The Race IAT shows a mean of 0.426 (SD=0.435), indicating a preference for whites over blacks. This finding is similar to prior studies (Amodio and Devine, 2006). For example, the race IAT reported in Lane et al. (2007) and available on a public website (https://implicit.harvard.edu/implicit/) reports a D score of 0.37 (n=732,881).

[Insert table 6 and 7 about here]

The self-report race attitude measure (ATB) shows a mean of 2.89 (SD=0.93), indicating a positive attitude towards blacks. The mean feeling thermometer (FT) score was M=0.72 (SD=1.58), indicating a preference for whites over blacks. Note that the ATB is an absolute measure of attitude towards blacks, whereas the FT is a relative measure. Although participants report a positive attitude towards blacks they also indicate a relative preference for whites over blacks.[17,18]

[17] One-way ANOVAs with Race IAT or ATB as the dependent variable and RACE as the independent variable indicate that Race IAT and ATB do not differ across conditions. However, an ANOVA with FT as the dependent variable and RACE as the independent variable does indicate differences across conditions (F=2.78, p=0.052). Post hoc Tukey HSD comparisons of FT means show two marginally significant differences: wb compared with bw (p=0.105) and wb compared with bb (p=0.065). Participants assigned to the wb race condition indicated the highest preference for whites over blacks (M=1.62).

[18] The feeling thermometer, along with the ATB and IAT, was administered after performance evaluations were completed and therefore participants had been exposed to the race manipulation. We did not manipulate order because prior research has found order does not moderate relations between implicitly and explicitly measured attitudes (Lane et al., 2007) or relations between behaviors and implicitly or explicitly measured attitudes (Greenwald et al., 2009a).

H3 states that implicitly measured attitudes predict race bias in performance evaluations. To test this hypothesis we conduct Pearson's product-moment correlations between the IAT measure and performance evaluation differences (DIFFOall) for each of the two mixed race conditions (wb and bw). Table 7 presents the correlation results. In the wb race condition there is a marginally significant correlation between IAT and DIFFOall (r=0.495, p=0.086). However in the bw race condition there is no correlation between IAT and DIFFOall (r=0.100, p=0.796). Therefore H3 is partially supported in that the expected correlation was found in the wb race condition but not the bw race condition.

Explicit racism is certainly not acceptable to most Americans today. Though for that reason we did not expect self-report measures of race attitude to predict race bias in performance evaluations, we did conduct Pearson's product-moment correlations between the two explicit race attitude measures (ATB and FT) and performance evaluation difference (DIFFOall) for participants in the wb and bw race conditions. ATB is not correlated with DIFFOall in either the wb race condition (r=0.059, p=0.847) or the bw race condition (r=0.-0.167, p=0.623). FT is not correlated with DIFFOall in either the wb race condition (r=0.382, p=0.198) or the bw race condition (r=0.-0.279, p=0.407). Consequently, as expected, explicit racial attitudes are not predictive of racial bias in performance evaluations.[19]We see two results of interest from H3 tests. The first result is that implicitly measured attitudes (IAT) predict performance evaluation differences between white and black managers, but self-report measures (ATB, FT) do not. This result has implications for addressing bias in accounting-based performance evaluation systems where different approaches may be needed to influence implicit versus explicit biases.

Also, the result is consistent with findings in other domains. For example, in a meta-analysis of the predictive ability of implicitly measured attitudes intergroup and interracial behaviors were more strongly related to implicit attitudes than to self-report measures (Greenwald et al., 2009a). Reporting racial attitudes on a self-report scale allows participants to control their responses in the context of a socially sensitive topic. Indeed, social sensitivity was found to moderate the effect of self-report measures on behavior significantly more than implicit attitudes (Greenwald et al., 2009a). The second result is that the IAT effect only emerged when a white manager outperformed a black manager, but not when a black manager outperformed a white manager. From a white (biased) perspective, a black outperforming a white may challenge the accepted order such that generally held implicit attitudes are overridden by

[19] A multiple regression with DIFFOall as the dependent variable and IAT and ATB as independent variables results in a significant coefficient for IAT (p=.040) and a non-significant coefficient for ATB (p=0.209). In this case ATB does not add incremental explanatory power when controlling for IAT. The same regression analysis with IAT and FT as independent variables results in a nonsignificant coefficient for IAT (p=0.251) and a nonsignificant coefficient for FT (p=0.699). However, with this regression there is a substantial collinearity concern with the VIF statistic equal to 4.334. Additionally, the IAT and FT measures were significantly correlated (r=0.877, p<0.001).

specific details of the performance evaluation task (Rudman and Lee, 2002). This finding deserves further investigation.

H4 states that implicitly measured race attitudes predict race bias in bonus allocations. To test this hypothesis we conduct Pearson's product-moment correlations between the IAT measure and bonus allocation (BonusA) for participants in the wb and bw race conditions. In the wb race condition there is a significant correlation between IAT and BonusA (r=0.761, p=0.003). However, in the bw race condition there is no correlation between IAT and BonusA (r=-0.152, p=0.656). Therefore H4 is partially supported in that the expected correlation was found in the wb condition but not in the bw race condition.

For the same reasons that we did not expect self-report measures of attitudes towards race to predict race bias in performance evaluation decisions, we did not expect self-report measures of race attitude to predict race bias in bonus allocations. We conduct Pearson's product-moment correlations between the two explicit race attitude measures (ATB and FT) and bonus allocation (BonusA) for participants in the wb and bw conditions. In the wb race condition there is a marginally significant correlation between ATB and BonusA (r=0.482, p=0.096), but there is no correlation between ATB and BonusA (r=-0.035, p=0.918) in the bw race condition. FT is significantly correlated with BonusA (r=0.580, p=0.038) in the wb race condition, but there is no correlation between FT and BonusA (r=-0.200, p=0.556) in the bw race condition. Therefore contrary to expectation in the wb race condition race attitude can predict bonus allocations depending on the explicit measure used, but it does not predict bonus allocations in the bw race condition.

The pattern of results for bonus allocations is similar to that of performance evaluation differences, with the addition that self-report measures were also predictive of bonus allocations when a white manager outperforms a black manager.[20]

5. Conclusions and implications

Overall, the results of this study indicate that accounting practices related to performance evaluation may be influenced by implicit attitudes that yield some degree of racial bias in decisions. At a practical level, neither racial bias nor the consequences of such bias are ever acceptable; prejudices like these are simply reprehensible. Yet, the fact that prejudicial attitudes are implicit or tacit phenomena

[20] Since race of managers does not differ in the ww and bb race conditions, no relations between race attitudes and performance evaluations or bonus allocation decisions were hypothesized. As anticipated the IAT, ATB, and FT measures for ww and bb are unrelated to performance evaluations and bonus allocations (all P-values >0.212, not reported in tables).

suggests that self-awareness may not be adequate to bring such prejudices to the level of consciousness for those who possess them. The force of the implicit within the ontology of selfhood is one example of just how complex and multidimensional human subjectivity is – our decisions and actions are influenced by personality factors embedded in cultural history in ways that do not rise to the level of consciousness for us. We should be vigilant in our awareness of and our efforts to eliminate the social reproduction of racial prejudice, particularly since it seems to operate in a tacit manner.

Accounting discourse and practices focused on performance evaluation are powerful forces over the construction of subjectivity, of human identity. Whether we are aware of any racial prejudices or not is less important than the fact that our evaluations do in fact construct identity – economic identity, moral identity, social identity, and organizational identity. The force of ethics is not located in our consciousness; it is instead interpretable and understandable in terms of what we are doing to others in the act of accounting for them. If racial prejudice is a component of that construction of identity, then we have strong grounds for moral pause in reflecting on what accounting does *to* persons.

For the most part, accounting research into performance evaluation has assumed rational and cognitive information processing models which facilitate inquiry into the relation between pieces of information and their effects on decisions (Hopwood, 2008). This study follows broad streams of IAT research to suggest that affective comportments of identity, even comportments about which we may be unaware, in this case implicit attitudes about race, are forces at work in performance evaluations of persons. As others have suggested (see for example Kennedy 1993, 1995) finding a way to mitigate bias is not a simple and obvious task. It is however a task of paramount practical, functional, and ethical importance.

This study raises questions of relevance to future research and practice. The fact that implicit racial bias seems conditioned by task specificity – whether one is simply evaluating someone or whether one is allocating bonus dollars – is both practically relevant and quite appealing from an experimental perspective. What is it about bonus allocation that causes it to seemingly render implicit racial bias more innocuous? Other interesting possibilities have to do with differences in outcomes as one moves across the four dimensions of the balanced scorecard; it seems that the force of implied racial attitude is more salient for the two "softer" BSC categories (customer related and learning and growth) than it is for the other two categories (financial and internal business processes).

The IAT and the scope of existing research suggests hundreds of ways in which the force of implicit attitudes is operative across a range of psychological, social, political, cultural, and ethical aspects of experience; the expanse of this research literature is represented well in Petty, Fazio, and Briñol (2009). Gender, sexual orientation, class-identity, ethnic-identity, religious-identity, and any number of categories of personhood which are historically painted with prejudicial and degrading force are

interesting domains for meaningful accounting inquiry, and the level of development of this area in social psychology provides plenty of guidance on how such accounting inquiry might proceed. Accounting has historically disclaimed any force other than "neutral" and "objective" measurement of others – it is time that that risible myth be exposed for what it is. IAT research can go a long way in yielding such exposure.

At a minimum, this study opens experimental possibilities for study of a richer, more psychologically and morally well-specified understanding of the relation between evaluators, accounting systems and processes, and those being evaluated. In addition to attitudes toward race, attitudes toward gender (see Gold et al., 2009), class, particular employee functions, and particular kinds of work may mean that performance evaluations, and other accounting-based decisions, are conditioned by a wide range of affective and largely unconscious comportments of identity and personality. Readers are encouraged to examine the hundreds of ways in which social-cognitive psychologists have deployed instruments like the IAT to move us closer to a phenomenologically rich understanding of just how complex evaluative judgments of others are. The fact that so very many people are judged routinely through accounting makes accounting a profoundly important area for this sort of research. The fact that the IAT is so well developed and so readily accessible means that there are few barriers to entry into this novel stream of accounting research. It is a particularly appealing approach for those interested in experimental studies aligned with the emancipatory interest that motivates critical-theoretic approaches to accounting. It is also an area where empirical research merges nicely with those critical-theoretic concerns for emancipation.

Beyond the experimental possibilities, the capacity of this area of research to enhance more philosophical and qualitative approaches is rich indeed. From the outset of this paper, we have motivated our interest in experimental study of implicit racial prejudice from the standpoint of the social ontology of subjectivity, of the construction of identity. Social ontological concerns are almost never addressed in quantitative research studies, though they are ubiquitous to critical work. In this paper, we have tried to produce <u>one</u> example of how quantitative research might accommodate social-ontological concerns. Nevertheless, because the most important consequence of racial prejudice is what it means for the socially constituted identity of its victims, critical, qualitative and humanistic research approaches are perhaps of <u>*more value*</u> in future research efforts than are the quantitative approaches that we have employed. But accounting researchers interested in social ontology and the construction of identity can deploy arguments drawn from this "third wave" of research into prejudices and biases to clearly link the discourses and practices of accounting to many aspects of identity – race, sex and sexual preference, class, ethnicity, types of labor, geography, etc. If, as many have held, Modernity is the study of the individual, then a major aspect of postmodern inquiry – including this area – has to do with understanding the force of

cultural practices and ideologies on what we have come to mean *by* the term "the individual." Exposing prejudicial ways in which Modernity has constituted identity is essential to overcoming the pretensions of modern accounting, pretensions like claims to fairness, neutrality, objectivity, progress, etc. We encourage critical and "qualitative" researchers to welcome the ways in which implied attitude testing can enhance this "overcoming" agenda for critical work. To do so, we must all overcome tendencies to approach scholarship as if it could be described with binaries like "qualitative" and "quantitative."

Racism and racial prejudice are not unitary concepts, nor is their relevance limited to individuals. Largely because of the modern focus on both clarity of terms and on individuals, "racism" has not been traditionally understood in a way that captures its complexity, nor has the racism inherent in institutions been examined enough.[21] Research agendas which focus on the complexity of racism (for example its embeddedness in other forms of oppression) can, like implied attitude studies, expose the "nonobvious," "noncognitive," somewhat invisible modes in which prejudice operates. In a study which engages in a more complex sense of racism and prejudice in the workplace than we address in this study, Fearfull and Kamenou (2006) provide an excellent example of how prejudice is embedded in institutions. That study is a model for those interested in more ethnographic approaches to the study of contemporary manifestations of prejudice in the workplace. Despite the differences in our approach and theirs, we share in common a concern to understand prejudice as more than confining it to the explicit racism of individuals can accommodate.

A strong limitation of this study has to do with the fact that we only examined the attitudes of white Americans; and, even within that category, a nonrepresentative sample of all white Americans. While much evidence indicates that groups other than white Americans demonstrate implicit bias against blacks (Nosek et al. 2007b), it is certainly wrong both practically and experimentally to make any inferences about other groups in the context of this study. Future research in this area should select samples from other populations.

Albert Einstein has been quoted as saying "[I]t is harder to crack a prejudice than an atom." And that makes it rather naïve to assume that prejudice in areas like performance evaluation can be addressed through appeals to conscious awareness and thereby remedied. Our implied attitudes are deeply engrained in tacit dimensions of our being, and attention to "the third wave" of research into prejudice can

---

[21] Institutional racism and its relevance to accounting and the workplace is expressed well in the following: *It is important to note that . . . an institution can be racist whether or not the individuals in that institution are racist or prejudiced. As an extension to the definition, we have taken the argument, well defined in the literature, that women in general are subject to sexist regimes within organisations. This leads, in career terms, to androcentricity (Crompton and Sanderson, 1990; Evetts, 1994; Halford et al., 1997), whereby women's careers are defined by male work patterns. Our data suggests strongly that organisations are, institutionally, both sexist and racist and that this combination of andro- and ethno- centricity has clear and potentially detrimental implications for ethnic minority women.* (Fearfull and Kamanou, 2006, p. 896).

help us understand that guarding against the force and consequences of prejudice requires vigilance, both cognitive and moral vigilance. This seems particularly important in areas like performance evaluation where the force of prejudice has real material consequences (e.g., promotions, compensation, dismissal) for those who are the objects of such evaluations. The ubiquity of accounting, its force over the lives of millions of workers, seems among the most meaningful domains of experience for critical inquiry into the full range of effects of prejudice – implicit, explicit, or both.

References

Amodio DM, Devine PG. Stereotyping and evaluation in implicit race bias: evidence for independent constructs and unique effects on behavior. Journal of Personality and Social Psychology 2006;91:652–61.

Atkinson AA, Balakrishnan R, Booth P, Cote JM, Groot T, Malmi T, Roberts H, Uliana E, Wu A. New directions in management accounting research. Journal of Management Accounting Research 1997;9:79-108.

Baiman S, Rajan MV. The informational advantages of discretionary bonus schemes. The Accounting Review 1995;70(4):557-79.

Banaji MR, Bazerman MH, Chugh D. How (un)ethical are you? Harvard Business Review 2003;81(12):56-64.

Banaji MR, Heiphetz L. Attitudes. In: Fiske ST, Gilbert DT, Lindzey G, editors. Handbook of social psychology (5th ed.). Hoboken, NJ: John Wiley & Sons, Inc; 2010. pp. 348-88.

Banker RD, Chang H, Pizzini MJ. The balanced scorecard: judgmental effects of performance measures linked to strategy. The Accounting Review 2004;79(1):1–23.

Baron AS, Banaji MR. The development of implicit attitudes: evidence of race evaluations from ages 6, 10, and adulthood. Psychological Science 2006:17:53–8.

Bertrand M. Mullainathan S. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. American Economic Review 2004;94(4):991–1013.

Bonner SE, Sprinkle GB. The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. Accounting, Organizations and Society 2002;27(4-5):303-45.

Brendl CM, Markman AB, Messner C. Indirectly measuring evaluations of several attitude objects in relation to a neutral reference point. Journal of Experimental Social Psychology 2005;41:346–68.

Brigham JC. College students' racial attitudes. Journal of Applied and Social Psychology 1993;23:1933–1967.

Campbell A. White attitudes toward black people. Ann Arbor, MI: Institute for Social Research, The University of Michigan; 1971.

Cardinaels E, van Veen-Dirks PMG. Financial versus non-financial information: the impact of information organization and presentation in a Balanced Scorecard. Accounting, Organizations and Society 2010;35(6):565-578.

Castilla EJ. Gender, race, and meritocracy in organizational careers. American Journal of Sociology 2008;113(6):1479–526.

Chugh D. Societal and managerial implications of implicit social cognition: Why milliseconds matter. Social Justice Research 2004;17:203–22.

De Houwer J. The extrinsic affective Simon task. Experimental Psychology 2003;50:77–85.

Devine PG, Plant EA, Amodio DM, Harmon-Jones E, Vance SL. The regulation of explicit and implicit race bias: the role of motivations to respond without prejudice. Journal of Personality and Social Psychology 2002;82:835–48.

Dilla WN, Steinbart PJ. Relative weighting of common and unique balanced scorecard measures by knowledgeable decision makers. Behavioral Research in Accounting 2005;17:43–53.

Dovidio JF. On the nature of contemporary prejudice: The third wave. Journal of Social Issues 2001;829-849.

Dovidio JF, Fazio R H. (1992). New technologies for the direct and indirect assessment of attitudes. In J. Tanur (Ed.), *Questions about survey questions: Meaning, memory, attitudes, and social interaction* (pp. 204–237). New York: Russell Sage Foundation.

Dovidio JF, Gaertner SL, Anastasio PA, Sanitioso R. Cognitive and motivational bases of bias: The implications of aversive racism for attitudes toward Hispanics. In S. Knouse, P. Rosenfeld, and A. Culbertson (Eds.), Hispanics in the Workplace. Newbury Park, CA: Sage, 1992, 75-106.

Dovidio JF, Kawakami K, Smoak N, Gaertner SL. The nature of contemporary racial prejudice: insight from implicit and explicit measures of attitudes. In: Petty RE, Fazio RH, Briñol P, editors. Attitudes: insights from the new implicit measures. New York, NY: Psychology Press; 2009. pp. 165-92.

Evetts J, editor. Women and career: themes and issues in advanced industrial societies. UK: Longman Group UK Limited; 1994.

Fazio RH. Multiple processes by which attitudes guide behavior: the MODE model as an integrative framework. In: Zanna MP, editor. Advances in experimental social psychology, 23. New York: Academic Press; 1990. pp. 75–109.

Fazio R, Jackson J, Dunton B, Williams C. Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? Journal of Personality and Social Psychology 1995;69:1013–27.

Fearfull A. Kamenou N. How do you account for it?: a critical exploration of career opportunities for and experiences of ethnic minority women. Critical Perspectives on Accounting 2006;17(7):883-901.

Feltham G, Xie J. Performance measure congruity and diversity in multi-task principal/agent relations. The Accounting Review 1994;69(3):429–54.

Gold A, Hunton JE, Gomaa MI. The impact of client and auditor gender on auditors' judgments. Accounting Horizons 2009;23(1):1-18.

Green AR, Carney DR, Pallin DJ, Ngo LH, Raymond KL, Iezzoni LI, Banaji MR. Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. Journal of General Internal Medicine 2007;22:1231–8.

Greenwald AG, Banaji MR. Implicit social cognition: attitudes, self-esteem, and stereotypes. Psychological Review 1995;102(1):4–27.

Greenwald AG, McGhee DE, Schwartz JLK. Measuring individual differences in implicit cognition: the Implicit Association Test. Journal of Personality and Social Psychology 1998: 74; 1464–80.

Greenwald AG, Nosek BA, Banaji MR. Understanding and using the Implicit Association Test: I. an improved scoring algorithm. Journal of Personality and Social Psychology 2003: 85; 197–216.

Greenwald AG, Krieger LH. Implicit bias: Scientific foundations. California Law Review 2006;94(4):945-67.

Greenwald AG, Poehlman TA, Uhlmann EL, Banaji MR. Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. Journal of Personality and Social Psychology 2009a;97:17-41.

Greenwald AG, Smith CT, Sriram N, Bar-Anam Y, Nosek BA. Implicit race attitudes predicted vote in the 2008 U.S. presidential election. Analyses of Social Issues and Public Policy 2009b;9(1):241-53.

Giuliano L, Levine DI, Leonard J. Manager race and the race of new hires. Journal of Labor Economics 2009;27(4):589-631.

Halford S, Savage M, Witz A. Gender, careers and organisations: current developments in banking, nursing and local government. UK: Macmillan Press Ltd; 1997.

Hebl MR, Foster JB, Mannix LM, Dovidio JF. Formal and interpersonal discrimination: A field study of bias toward homosexual applicants. Personality and Social Psychology Bulletin 2002;28:815-25.

Hofmann W, Gawronski B, Gschwendner T, Le H, Schmitt M. A meta-analysis on the correlation between the implicit association test and explicit self-report measures. Personality and Social Psychology Bulletin 2005;31:1369–85.

Hopwood AG. Management accounting research in a changing world. Journal of Management Accounting Research 2008;20:3-13.

Hovland CI, Janis Il, Kelley HH. Communication and persuasion: psychological studies of opinion change. New Haven, CT: Yale University Press; 1953.

Humphreys KA, Trotman KT. The balanced scorecard: the effect of strategy information on performance evaluation judgments. Australian School of Business Research Paper No. 2010 ACCT 02; 2010.
Ittner CD, Larcker DF. Innovations in performance measurement: trends and research implications. Journal of Management Accounting Research 1998;10:205–38.
Ittner CD, Larcker DF. Assessing empirical research in managerial accounting: a value-based management perspective. Journal of Accounting and Economics 2001;32:349–410.
Ittner CD, Larcker DF. Coming up short on nonfinancial performance measurement. Harvard Business Review 2003;81(11):88-95.
Ittner CD, Larcker DF, Meyer MW. Subjectivity and the weighting of performance measures: evidence from a balanced scorecard. The Accounting Review 2003;78(3):725–58.
Johnson EN, Lowe DJ, Reckers PMJ. Alternative work arrangements and perceived career success: current evidence from the big four firms in the US. Accounting, Organizations and Society 2008;33:48–72.
Kaplan RS, Norton DP. The balanced scorecard – measures that drive performance. Harvard Business Review 1992;70(1):71–9.
Kaplan RS, Norton DP. The balanced scorecard: translating strategy into action. Boston, MA: Harvard Business School Press; 1996.
Kaplan RS, Norton DP. Having trouble with your strategy? Then map it. Harvard Business Review 2000;78(5):167-76.
Kaplan SE, Petersen MJ, Samuels JA. Effects of subordinate likeability and balanced scorecard format on performance-related judgments. Advances in Accounting 2008;23:85–111.
Kennedy J. Debiasing audit judgment with accountability: a framework and experimental results. Journal of Accounting Research 1993;31(2):231-45.
Kennedy, J. Debiasing the curse of knowledge in audit judgment. The Accounting Review 1995;70(2):247-73.
Keppel G. Design and analysis: a researcher's handbook. Upper Saddle River, NJ: Prentice Hall; 1991.
Kraus SJ. Attitudes and the prediction of behavior: a meta-analysis of the empirical literature. Personality and Social Psychology Bulletin 1995;21:58–75.
Kraiger K, Ford JK. A Meta-Analysis of Ratee Race Effects in Performance Ratings. Journal of Applied Psychology 1985;70(1):56-65.
Lane KA, Banaji MR, Nosek BA, Greenwald AG. Understanding and using the implicit association test: IV. What we know (so far) about the method. In: Wittenbrink B, Schwarz N, editors. Implicit measures of attitudes. New York, NY: The Guilford Press; 2007. pp. 59-102.
Libby T, Salterio SE, Webb A. The balanced scorecard: the effects of assurance and process accountability on managerial judgment. The Accounting Review 2004;79(4):1075–94.
Liedtka SL, Church BK, Ray MR. Performance variability, ambiguity intolerance, and balanced scorecard-based performance assessments. Behavioral Research in Accounting 2008;20(2):73-88.
Lipe MG, Salterio SE. The balanced scorecard: judgmental effects of common and unique performance measures. The Accounting Review 2000;75(3):283–98.
Lipe MG, Salterio S. A note on the judgmental effects of the balanced scorecard's information organization. Accounting, Organization and Society 2002;27(6):531–40.
Luft J. Nonfinancial information and accounting: a reconsideration of benefits and challenges. Accounting Horizons 2009;23(3):307–25.
Luo, M. In job hunt, college degree can't close racial gap. The New York Times 2009: December 1.
Malina M, Selto F. Communicating and controlling strategy: An empirical study of the effectiveness of the balanced scorecard. Journal of Management Accounting Research 2001;13:47–90.
McConnell AR, Leibold JM. Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. Journal of Experimental Social Psychology 2001;37:435–42.
McKay PF, McDaniel MA. A reexamination of black–white mean differences in work performance: more data, more moderators. Journal of Applied Psychology 2006;91(3):538–54.

Merchant KA. Measuring general managers' performances: market, accounting and combination-of-measures systems. Accounting, Auditing & Accountability Journal 2006;19(6):893-917.
Messick D, Bazerman M. Ethical leadership and the psychology of decision making. Sloan Management Review 1996;37(2):9-22.
Meyer DE, Schvaneveldt RW. Facilitation in recognizing pairs of words: Evidence of dependence between retrieval operations. Journal of Experimental Psychology 1971;90:227-34.
Miller P, O'Leary T. Accounting and the construction of the governable person. Accounting, Organizations and Society 1987;12(3):235-65.
Moers F. Discretion and bias in performance evaluation: the impact of diversity and subjectivity. Accounting, Organizations and Society 2005;30:67–80.
Nørreklit H. The balance on the balanced scorecard – a critical analysis of some of its assumptions. Management Accounting Research 2000;11:65-88.
Nosek BA, Banaji MR. The go/no-go association task. Social Cognition 2001;19:625–66.
Nosek BA. Moderators of the relationship between implicit and explicit evaluation. Journal of Experimental Psychology: General 2005;135:565–84.
Nosek BA, Greenwald AG, Banaji MR. The implicit association test at age 7: a methodological and conceptual review. In: Bargh JA, editor. Automatic processes in social thinking and behavior. New York, NY: Psychology Press; 2007a. pp. 265-92.
Nosek BA, Smyth FL, Hansen JJ, Devos T, Lindner NM, Ranganath KA, Smith CT, Olson KR, Chugh D, Greenwald AG, Banaji MR. Pervasiveness and correlates of implicit attitudes and stereotypes. European Review of Social Psychology 2007b;18:36 – 88.
Nosek BA, Hansen J. Personalizing the Implicit Association Test increases explicit evaluation of target concepts. European Journal of Personality Assessment 2008;25:226–36.
Payne BK, Cheng CM, Govorun O, Stewart BD. An inkblot for attitudes: affect misattribution as implicit measurement. Journal of Personality and Social Psychology 2005;89:277–93.
Payne BK, Burkley MA, Stokes MB. Why do implicit and explicit attitude tests diverge? The role of structural fit. Journal of Personality and Social Psychology 2008;94(1):16–31.
Peery D, Bodenhausen G. Ambiguity and ambivalence in the voting booth and beyond: a social-psychological perspective on racial attitudes and behavior in the Obama era. Du Bois Review: Social Science Research on Race 2009;6:71-82.
Petersen T, Saporta I. The opportunity structure for discrimination. American Journal of Sociology 2004;109(4):852–902.
Petty RE, Fazio RH, Briñol P. The new implicit measures: an overview. In: Petty RE, Fazio RH, Briñol P, editors. Attitudes: insights from the new implicit measures. New York, NY: Psychology Press; 2009. pp. 3-18.
Power M, Laughlin R. Habermas, law and accounting. Accounting, Organizations and Society 1996;21(5):441-65.
Pulakos ED, Oppler SH, White LA, Borman WC. Examination of race and sex effects on performance ratings. Journal of Applied Psychology 1989;74(5):770-80.
Rigby D, Bilodeau B. Management tools and trends 2009. Boston, MA: Bain & Company; 2009.
Roberts J. The possibilities of accountability. Accounting, Organizations and Society 1991;16(4):355-68.
Roberts J. No one is perfect: the limits of transparency and an ethic for 'intellligent' accountability. Accounting, Organizations and Society 2009;34(8):957-70.
Roberts ML, Albright TL, Hibbets AR. Debiasing balanced scorecard evaluations. Behavioral Research in Accounting 2004;16:75–88.
Roth PL, Huffcutt AI, Bobko P. Ethnic group differences in measures of job performance: a new meta-analysis. Journal of Applied Psychology 2003;88(4):694–706.
Rudman LA, Lee MR. Implicit and explicit consequences of exposure to violent and misogynous rap music. Group Process & Intergroup Relations 2002;5:133–50.
Rudman LA. Social justice in our minds, homes, and society: the nature, causes, and consequences of implicit bias. Social Justice Research 2004;17:129–42.

Scott TW, Tiessen P. Performance measurement and managerial teams. Accounting, Organizations and Society 1999;24:263-85.

Shearer T, Arrington C. Accounting in other wor[l]ds: a feminism without reserve. Accounting, Organizations and Society 1993;18(2 and 3):253-72.

Tayler WB. The balanced scorecard as a strategy-evaluation tool: the effects of implementation involvement and a causal-chain focus. The Accounting Review 2010;85(3):1095-117.

Waldman DA, Avolio BJ. Race Effects in performance evaluations: controlling for ability, education, and experience. Journal of Applied Psychology 1991;76(6):897-901.

Weisenfeld LW, Robinson-Backmon IB. Upward mobility and the African American accountant: an analysis of perceived discrimination, perceived career advancement curtailment, and intent to remain. Accounting and the Public Interest 2007;7:26-49.

West C. Race matters (Second edition) 2001. New York: Vintage Books.

Wilson T, Lindsey S, Schooler TY. A model of dual attitudes. Psychological Review 2000;107:101–26.

Wong-On-Wing B, Guo L, Li W, Yang D. Reducing conflict in balanced scorecard evaluations. Accounting, Organizations and Society 2007;32(4–5):363–77.