This is to certify that the

thesis entitled

Race Effects on Performance Evaluation in a Team Situation: An Attributional Perspective

presented by

Aleksander P.J. Ellis

has been accepted towards fulfillment
of the requirements for

M.A. degree in Psychology

Major professor

Date 10/25/99

RACE EFFECTS ON PERFORMANCE EVALUATION IN A TEAM SITUATION: AN ATTRIBUTIONAL PERSPECTIVE

By

Aleksander P.J. Ellis

A THESIS

Submitted to
Michigan State University
In partial fulfillment of the requirements
For the degree of

MASTER OF ARTS

Department of Psychology

1999

ABSTRACT

RACE EFFECTS ON PERFORMANCE EVALUATION IN A TEAM SITUATION: AN ATTRIBUTIONAL PERSPECTIVE

By

Aleksander P.J. Ellis

This study examined the way in which the race of a team leader influenced attributions and evaluations of leader performance on a team task with distributed expertise. There were two main purposes of this research: to introduce attribution theory as an explanation of racial differences in performance rating and to examine performance ratings in a team situation, where subordinates (the raters) interacted with their superior (the ratee). Research on performance rating and race has, for the most part, ignored both attribution theory and team settings when attempting to explain racial differences in evaluating an individual's performance level. This study focused solely on racial differences in ratings of White American versus African American team leaders. It was proposed that certain attributions based on racial stereotypes would be created depending on the team leader's level of performance. These attributions would then moderate the direct relationship between a team leader's level of performance and performance ratings. In addition, it was hypothesized that race would also have a direct influence on the level of performance ratings. The results did not support the original model, since race was not directly related to performance ratings. Instead, it was found that race interacted with the team leader's actual performance in predicting performance ratings. In the high performance condition, attributions mediated the relationship between race and performance ratings. The practical and theoretical implications of these results are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Over past years there has been a considerable amount of research in the area of performance appraisal, which has led investigators to look at individual characteristics in explaining variance in performance ratings (Landy & Farr, 1980). One focus has been on the characteristic of race. Several early studies investigating the effects of race on performance ratings found that raters tended to give higher ratings to those of the same race (Cox & Krumboltz, 1958; DeJung & Kaplan, 1962). Further experimentation supported these findings by showing that subjects favored applicants of the same race even when objective measures were defined (Hamner, Kim, Baird, & Bigoness, 1974). Yet several studies disagreed with the assertion that African Americans rate African Americans higher and White Americans rate White Americans higher. For instance, it was discovered that when performance was high, African Americans and White Americans were rated similarly, and when performance was low, African Americans were actually rated higher than White Americans (Bigoness 1976; Rotter & Rotter 1969). It was then proposed that the proportion of racially similar individuals in an assessment group would have a significant effect on ratings. Although the effects were small, White American males tended to be rated higher, while African American females tended to be rated lower, when the proportion of White American males in the assessment group increased (Schmitt & Hill, 1977). There were also several studies that found no significant race effects on performance ratings (e.g. Bass & Turner, 1973; Fox & Lefkowitz, 1974; Schmidt & Johnson, 1973).

With the apparent lack of consistent effects on ratings in many of the aforementioned field studies (e.g. Bass & Turner, 1973), Wendelken and Inn (1981)

argued that the effects found in the early laboratory studies were not generalizable to the organizational environment. Face-to-face interaction within the workplace, according to the authors, would dampen the effect of group stereotypes directed at individual group members. Their field data supported this assertion by showing that insignificant proportions of the variance in performance ratings were accounted for by the race of the rater, the individual being rated, and their interaction. Others have argued to the contrary; their position was that the effects should be less likely in the laboratory because of college students' inclination to provide socially desirable responses (Jones & Signall, 1971). In summary, a decisive conclusion with respect to racial bias in performance ratings could not be reached from the early literature.

Kraiger and Ford (1985) attempted to consolidate these diverse findings across settings and populations by conducting a meta-analysis in order to obtain more definitive data on the effect of race on performance evaluations. Their review included over seventy studies, most of which came directly from field samples. The results yielded a mean point-biserial correlation between the race of the ratee and performance ratings of .18 (corrected for unreliability) for White American raters and -.22 for African American raters. Respectively, these two correlations determined 1.0% and 3.2% of the population variance for White American and African American raters. In other words, African Americans gave higher ratings to African Americans and White Americans gave higher ratings to White Americans. The authors concluded that race has a moderate impact on performance ratings with both laboratory and field samples, directly contradicting the previously discussed results obtained by Wendelken and Inn (1981) and others (e.g. Bass & Turner, 1973).

However, the meta-analytic results of Kraiger and Ford still did not allow for determining whether differences in ratings were due to bias or actual differences in performance (Kraiger & Ford, 1985; Hunter & Hirsch, 1987). Another problem was the lack of independence of the rating observations. As in most field research on performance rating measures, one rater evaluated more than one individual. This inevitably complicated the distinction between within-rater and between-rater variance (Pulakos, White, Oppler, & Borman, 1989).

Subsequent research by Pulakos et al. (1989) was directed at eliminating previous problems when examining the role of race in performance evaluations. The data were collected as part of Project A, a large scale, long term research endeavor conducted in the military to improve selection and classification of new army recruits (Eaton, Goer, Harris, & Zook, 1984). Data sets were formed where each individual was rated by two raters- one from the same racial group as the ratee and one from another. In this design, actual performance was held constant as well as the number of raters assigned to a particular individual. The results were slightly different from those of Kraiger and Ford (1985). In this case, race accounted for no more than 1% of the variance in ratings.

However, there were problems with the research done by Pulakos et al. (1989) that could have shrunk the observed effects by a significant proportion. The individuals who participated in the study also participated in a training program specifically designed to eliminate nonperformance factors (e.g. race) from performance ratings. In addition, the study was based entirely on a sample of military personnel. The percentage of African Americans in the military was, and still is, much higher than in most organizations

(Pulakos et al., 1989). Perhaps race was not as salient to those participating in the military study, and thus did not influence their ratings.

Sackett and DuBois (1991) attempted to translate the findings of Pulakos et al. (1989) to the civilian sector by using data from firms across the United States and combining it with the military data from Project A in order to expand the data base on rater-ratee race effects. Similar to previous research, the civilian data employed a design in which each individual was rated by two raters (i.e. one White American and one African American) allowing a within subjects analysis. The results obtained from the military and civilian data were strikingly different from those obtained by Kraiger and Ford (1985) and Pulakos et al. (1989). In both of those meta-analyses, African Americans rated African Americans higher and White Americans rated White Americans higher. Yet that conclusion was not supported by the extremely large amount of data collected by Sackett and DuBois. In fact, both African Americans and White Americans gave lower ratings to African Americans and higher ratings to White Americans in both low and high performance situations.

Although the research by both Pulakos et al. (1989) and Sackett and DuBois (1991) focused on solving methodological issues of previous research, they neglected the greater issue of understanding the cognitive processes underlying race effects on performance ratings. In support of this proposition, Kraiger & Ford (1985) noted that further research should concentrate on the cognitive processes involved in evaluating performance instead of the magnitude of racial bias.

Theories of the Cognitive Processes Underlying Racial Differences in Ratings

Several explanations have been proposed in the literature. For instance, Kraiger and Ford (1985) suggested that tokenism (Kanter, 1977) may explain why studies have found that, as the proportion of group members possessing a certain characteristic (e.g. race) decreases, that characteristic becomes easier to process and recall when searching for information about the individual (Taylor, 1981). These minority characteristics (e.g. race) "carry assumptions about culture, status and behavior highly salient for majority group members" (Kanter, 1977, p.968). According to Kanter, the "token" member is singled out and viewed negatively by the majority. A group member is seen as "token" when the subgroup to which he or she belongs comprises 35% or less of the work group. When 15% or less are present, the situation is referred to as "skewed," while more than 15% results in a "tilted" environment (Kanter, 1977). Research on tokenism applied to race has found mixed results. Although Ziller, Behringer, and Goodchilds (1960) found that when African Americans enter a group they tend to be characterized by the single characteristic of race, their results failed to reach statistical significance. However, similar studies have shown that for most comparisons, the racially distinctive member of a group was judged significantly less positively than the majority members (e.g. Craig, 1996). Firm conclusions on the theory of tokenism applied to performance ratings, unfortunately, are difficult to find due to the nonsignificant findings of other researchers. Sackett, DuBois, and Noe (1991) demonstrated that, when the proportion of African Americans in a work group increases, there is no significant difference in performance ratings, indicating an absence of group composition effects. Support for the theory becomes even more difficult to find when the data of Sackett and DuBois (1991) is taken

into account. Tokenism cannot explain why African Americans rated other African Americans lower than White Americans.

Aside from tokenism, other theories have also been applied to the issue of rating individuals based on race. Two theories, the similar-to-me effect and social identity theory, have been examined and seem to converge in their apparent explanation of the process underlying the evaluation of others. The crux of the similar-to-me effect is interpersonal attraction, where individual's are attracted to those similar to them (Byrne, 1971). And the result is often that higher ratings are given to similar individuals both in field (Lin, Dobbins, & Farth, 1992) and in laboratory (e.g. Rand & Wexley, 1975) studies.

Social identity theory posits that people inherently wish to sustain a positive self-regard, which leads them to view others in their same group as positive (Messick & Mackie, 1989, p.59). As a consequence, those individuals within the group are evaluated more favorably than outsiders (Tajfel, 1981). To become a member of such a group, the individual must possess certain characteristics such as belonging to the same social category (Turner, 1984, p.530).

In essence, both social identity theory and the similar-to-me effect agree that individuals of the same group (i.e. race) will rate each other in a more positive manner than those outside the group. This effect will take place only when other members of the group are present (Doise, 1978; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). In the case where the individual is the sole representative from his or her group, the person would instead opt to identify with the members of the majority (Davis et al., 1995, p.131). Support has been partially received for this theory. It was found that, in an

interview situation with equal African American and White American panel members, individuals tended to rate same-race job candidates higher. In addition, when an African American rater was alone among a group of White Americans, he or she tended to give higher ratings to the opposite race, supporting the contention that the similar-to-me effect only takes place with others of the same group. However, although the pattern was similar for the case with one White American interviewer, the results did not reach statistical significance and thus did not provide strong support for the argument that one White American will identify with the African American majority. (Prewett-Livingston, Field, Veres, & Lewis, 1996)

Other interview rating researchers have found no support for the similar-to-me effect and social identity theory (e.g. Campion, Pursell, & Brown, 1988). In other words, African Americans and White Americans were rated similarly by the interview panel. Unfortunately, although the studies mentioned above examined race, they based their ratings on interviews rather than observations of on-the-job performance. On the other hand, studies that did use on-the-job performance indices did not examine race (e.g. Cardy & Dobbins, 1986; Dovido & Gaertner, 1983). More importantly, none of the studies actually measured perceived similarity to the individual being rated.

Kraiger and Ford (1990) continued the search for the cognitive processes underlying racial bias in performance ratings by completing a second meta-analysis. They examined another two theories, intergroup theory (Alderfer, 1986) and positivity bias (Pettigrew, 1979), as possible explanations for their results. Intergroup theory proposes that evaluations of performance are influenced by actual performance level differences, but are also dependent on the perceptions of that performance. For example, if two

employees performed equally well, one would be perceived as having performed better because the rater identified with his or her race. The second theory, positivity bias, hypothesizes that those in the majority group will bolster the ratings of other majority group members. In this case, higher ratings are given to those of the majority race instead of lower ratings being given to the minority group member. The inflation of ratings is due to the addition of job-irrelevant factors when rating majority group members. Kraiger and Ford (1990) found that majority members (White Americans) tended to focus more on work performance (job-relevant) measures when rating minority members (African Americans), but did not focus more on job-knowledge (job-relevant) measures. In other words, firm support for the use of intergroup theory in understanding the process of evaluating performance was not evident in Kraiger and Ford's (1990) data. And both intergroup theory and positivity bias are unable to accurately explain why African Americans would rate other African Americans lower, as Sackett and DuBois (1991) previously reported.

In sum, tokenism, the similar-to-me effect, social identity theory, intergroup theory, and positivity bias received only mixed support from the literature on race effects in performance appraisal. More importantly, none of the research supporting these theories utilized the technique of meta-analysis to the extent that it was used by Sackett and DuBois (1991). Their civilian data included 12,022 White Americans rated by White American raters, 661 White Americans rated by African American raters, 5,972 African Americans rated by White American raters, and 1,110 African Americans rated by African American raters for the between subjects analyses. The within subjects analysis, where each individual was rated by both a African American and a White American rater,

consisted of 286 White Americans and 331 African Americans. The military study, done by Pulakos et al. (1989), used more than 35,000 individuals in the between subjects analysis and more than 1,800 individuals in the within subjects analysis. The sum of these two massive data sets by Sackett and DuBois created a very powerful tool with which to examine race effects in the process of performance evaluation. The civilian ratings of overall performance and part of the military results clearly showed that African Americans were rated lower not only by White Americans, but also by other African Americans. This result was in direct conflict with previous meta-analytic results (e.g. Kraiger & Ford, 1985) that African Americans and White Americans rate the same race higher. Another problem this created was that none of the theories previously enumerated could explain the psychological process behind giving lower ratings to individuals of the same race, especially since it was often a one on one rating system devoid of any group (i.e. token) effects. However, it is suggested here that one theory that has not been examined, attribution theory, could potentially explain variance in the relationship between race, actual performance, and performance ratings. The next section addresses it.

Attribution Theory

Bernard Weiner (1985) wrote that, when examining the behavior of another individual, one has an inherent need to know why someone did what they did. Was it due to ability, hard work, luck, or good looks? "The attributor is not simply an attributor, a seeker after knowledge; his latent goal in attaining knowledge is that of effective management of himself and his environment" (Kelly, 1971, p.22). The search for causality has been conducted in a wide variety of research contexts over many decades (e.g. Feather & Davenport, 1981; Heider, 1958; Rosenbaum, 1972; Rotter, 1966).

Weiner (1985), in reviewing many of the studies, concluded that past investigation had shown that there are only a small number of applicable causes that individuals utilize, and these causes all share three main properties: locus, stability, and controllability. The first property, locus, is used to assign cause either internally (due to the person) or externally (due to his or her environment). For example, cognitive ability is an internal cause while the difficulty of the task is an external cause, since it is not within the person. As for the second property, cognitive ability is also viewed as a stable cause while the difficulty of the task is viewed as unstable. Cognitive ability, in reference to the third property, is seen as uncontrollable since the individual does not have the option of immediately increasing their cognitive abilities.

Following Schachter and Singer's (1962) famous work dealing with emotional states, Weiner (1985) developed his own emotive theory. Experience, similar to Schachter and Singer's model, is refined and differentiated as more complex cognitions enter the emotive process (see Figure 1).

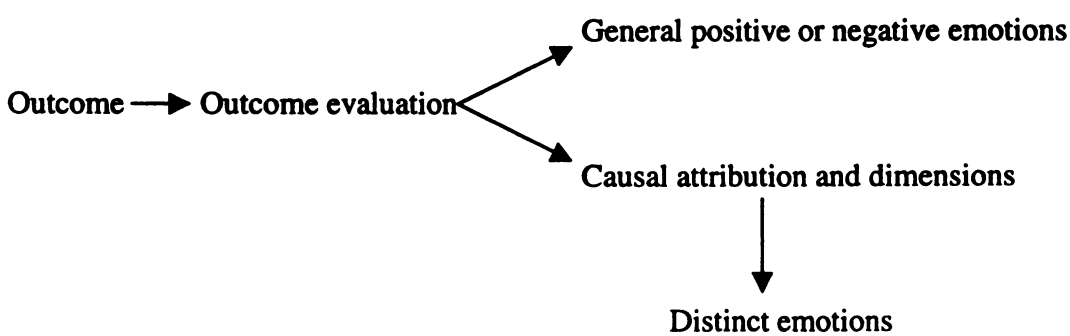

Figure 1. *Weiner's emotive theory* (from Weiner, 1985, p.560)

The first step in the abbreviated model above is a "primitive" one for the individual. A general negative or positive reaction occurs depending on whether the event was a success or a failure. This general reaction is attribution independent and outcome dependent. Following the immediate reaction, a reappraisal of the event takes place and the individual begins to search for causal attribution. This step is attribution dependent and, after the individual locates the appropriate causal ascription, different emotions will be formed according to the locus, controllability, and stability of the cause.

Attribution Theory and Performance Ratings

Weiner's (1985) emotive theory, as described above, can be directly relevant to the organizational environment. Often employees are asked to rate the performance of others within the workplace. When observing an individual performing, a general reaction takes place depending on whether their effort was a success or failure. Obviously individuals will be evaluated higher if their behavior was a success than if it was a failure. However, according to attribution theory, this evaluation is then refined as the observer searches for the appropriate causal attribution based on the locus, controllability, and stability of the cause. According to the theory, in the case of success, the best ratings will be given to the individual who is believed to have performed well because of some internal attribute such as ability, while those who performed well due to external attributes such as luck will be rated lower. In the case of failure, the best ratings will be given to the individual who performed poorly because of external attributes, while lower ratings will be given to those who performed poorly due to internal attributes.

Attribution Theory, Race, and Performance Ratings

Attribution theory, as described above, has not been used in the literature as a potential explanation for the effects of race on performance ratings. However, researchers have suggested that "good performance for minorities might be attributed to external factors (e.g., luck or the task) or unstable factors (e.g., effort) rather than to ability. Good performance for majority group members, on the other hand, might be more likely attributed to ability, which is both internal and stable" (Ilgen & Youtz, 1986, pp. 313-314).

It has been shown that African Americans, out of all the racial minorities in the United States, are more negatively stereotyped than anyone else (Smith, 1990). Attribution researchers propose that stereotypes like this create expectations concerning individual behavior and thus influence causal attributions (e.g., Deaux, 1976; Jones & McGillis, 1976). Behavior that is inconsistent with the stereotype is attributed to external causes, while consistencies are attributed to internal causes (e.g., Heilman, 1983; Nieva & Gutek, 1980). The research linking stereotypes and attributions has, for the most part, supported the assertion that the success of women and African Americans is more likely to be attributed to external causes such as luck than that of White males. Failure, however, is attributed more to internal causes such as ability (e.g., Ilgen & Youtz, 1986; O'Leary & Hansen, 1984). These results have important implications because, when the performer attributes his or her failure to the difficulty of the task (external-unstable), individuals will continue to work towards success. However, attributions of low ability will tend to lead to feelings of hopelessness in the face of future difficulty (Andrews & Debus, 1978; Chapin & Dyck, 1976; Dweck, 1975).

Although the literature appears consistent with an attribution interpretation, there is research that does not completely agree with the majority position. Jackson, Sullivan, and Hodge (1993) found that, in reviewing college applications of African Americans and White Americans, subjects viewed effort as an internal-stable cause when behavior was stereotype-inconsistent for a period of one year. Although part of their results differed slightly from previous research, Jackson and her colleagues' reached conclusions similar to the standard position that stereotypes influenced attributions, stereotype-inconsistent behavior (high performance for African Americans and low performance for White Americans) resulted in external attributions, stereotype-consistent behavior resulted in internal attributions, and attributions influenced evaluations.

However, the findings of Jackson et al. (1993), like many of the studies in the literature, may not relate to the performance rating contexts within organizations. For example, evaluations given to both African Americans and White Americans resulted from the examination of paper descriptions, which listed an individual's background and personal characteristics (e.g., Whitehead, Smith, & Eichhorn, 1982). Within an organization, however, there are face to face interactions between individuals that could influence performance ratings through a different process and could result in different conclusions. In fact, Pettigrew (1979) felt that a true test of attributional error could occur only when individuals interact. The present study was designed to closely mimic an actual organizational environment. Perhaps the communication between real people would increase the rating effects found using "paper people." More importantly, most studies in the literature do not consist of subjects actually observing an individual's performance firsthand. When subjects participate in research and end up looking at a

piece of paper, there is no personal connection to those they are rating. Finally, there are no consequences for the failure or success of the ratee in much of the literature. If subjects are working with an individual toward a common goal (e.g. a monetary reward), they may be more likely to rate that individual as an employee would in a real organization. In conclusion, face-to-face interaction, observing an individual's performance firsthand, and working toward a common goal are all critical boundary conditions that are lacking in most research to date.

DeCarlo and Leigh (1996) attempted to expand the attribution literature to the organization by looking at the impact of a salesperson's task and social attraction on a sales manager's causal attributions. The authors defined task attraction as "the desirability of the salesperson to the sales manager as a work partner" and social attraction as "the desirability of the salesperson to the sales manager as a friend and social partner" (DeCarlo & Leigh, 1996, p.48). Their basic model is pictured in Figure 2.

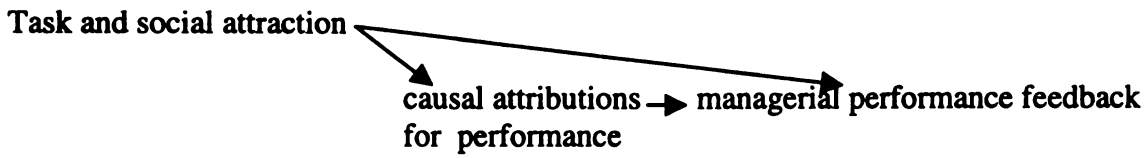

Figure 2. *Mediation model of the relationship between task and social attraction and performance feedback*

Causal attributions enter the picture as a partial mediator between task/social attraction and feedback. In other words, task and social attraction were hypothesized to directly affect feedback and to influence causal attributions which, in turn, were hypothesized to influence managerial performance feedback. Regarding the first link, those salespeople

high in social and task attraction benefited by receiving external causal attributions from their managers when they performed poorly and internal causal attributions when they performed well. On the other hand, those salespeople low in task and social attraction received internal attributions for poor performance and external attributions for good performance. Unfortunately, DeCarlo and Leigh did not include actual performance or performance ratings in their model. However, the influence of task and social attraction on causal attributions can be translated to the performance rating context.

Logically, the next step is to combine the previous research on attribution theory and the literature on race and performance ratings. In doing so, DeCarlo and Leigh's (1996) model can be used by replacing task and social attraction with racial stereotypes, which is supported by a number of researchers (e.g., Ilgen & Youtz, 1986; O'Leary & Hansen, 1984). It is expected that stereotypes concerning African Americans will function in much the same way as low task and social attraction. White American stereotypes, on the other hand, are hypothesized to coincide with high task and social attraction (see *Figure 3).*

Figure 3. *The relationship between race and causal attributions*

Bringing in the work of Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997), race is expected to have a direct effect on performance ratings (see

Figure 4). In their meta-analysis, Schmitt et al. found that race was correlated .18 with performance ratings.

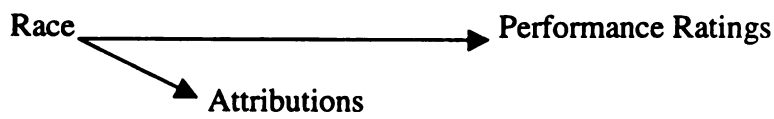

Figure 4. *The relationship between race and performance ratings*

Hamner, Kim, Baird, and Bigoness'(1974) research can also be added to the model, since they found that actual performance has a direct effect on performance ratings (see Figure 5). Poor performance resulted in low performance ratings, while good performance resulted in high performance ratings.

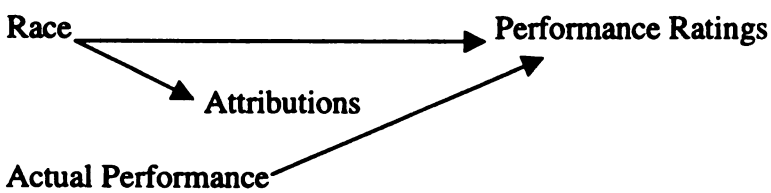

Figure 5. *The relationship between actual performance and performance ratings*

However, the direct effects of race on attributions clearly depend on the actual performance of an individual on a specific task, which is a moderated relationship (see Figure 6). Although there has been little relevant research on this link, some studies have indicated that racial effects on performance attributions are moderated by actual

performance (Whitehead, Smith, and Eichhorn, 1982). Ilgen & Youtz (1986) hypothesized that, based on previous research (e.g., Deaux & Emswiler, 1974; Heilman, 1983), African Americans would receive internal attributions for failure and external attributions for success while White Americans would receive external attributions for failure and internal attributions for success.

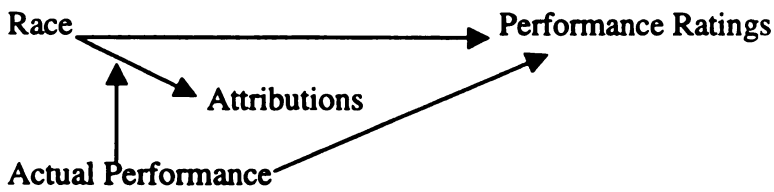

Figure 6. *The interaction between race and actual performance*

The last step is to connect causal attributions to performance ratings (see Figure 7). In other words, this study adds causal attributions to the model of race, actual performance, and performance ratings previously investiagted by meta-analytic researchers (e.g., Sackett & DuBois, 1991) in order to explain the cognitive processes underlying racial effects on performance ratings. It is proposed that causal attributions, based on an individual's race and actual performance, moderate the relationship between actual performance and performance ratings. Internal attributions for success will result in the highest performance ratings, with external attributions for success resulting in significantly lower ratings. External attributions for failure will result in even lower ratings and internal attributions will result in the lowest performance ratings.

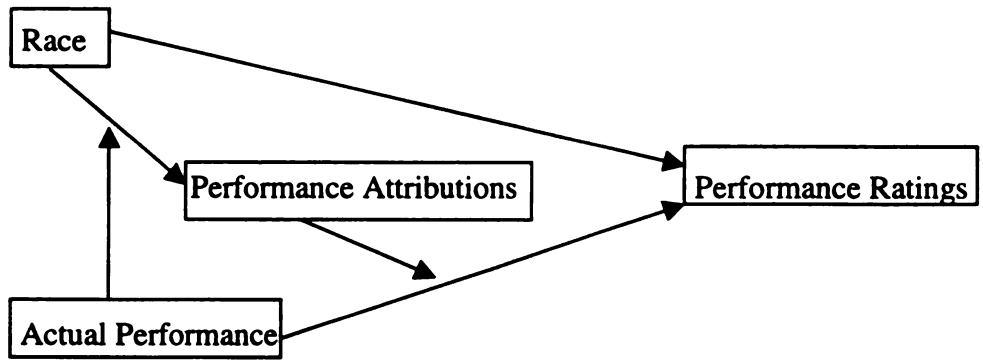

Figure 7. *The interaction between actual performance and causal attributions*

The above model was constructed using studies that examined ratings by one individual, usually a supervisor, of another individual, usually a subordinate. However, this model can also be applied to team member (i.e., subordinate) ratings of a team leader (i.e., supervisor). This study introduces teams to the literature on race and performance ratings for three reasons. First, previous laboratory research concentrated on ratings of "paper" African Americans and White Americans. These stimuli do not offer any opportunities for the rater to interact with the ratee as employees would within an organization. However, if rater and ratee are both working together on a collaborative task (i.e., as part of a team) they must have at least limited contact with each other. Second, an individual's performance within an organization influences other employees, including supervisors, subordinates, and peers. Lab research has, for the most part, focused on "paper people," where negative or positive performance is of no consequence to the rater. In a team situation, all participants are trying to reach a common goal (e.g., a

monetary reward) and therefore one team member's performance impacts the performance of the other team members, which could affect the performance evaluation process. Third, organizations are currently interested in gathering performance feedback not only from supervisors and peers, but also from subordinates (e.g., 360^0 feedback). So, in order for lab work to simulate the organizational environment, bias in performance ratings should be investigated from all angles. In other words, racial differences in performance ratings should be examined using subordinate raters as well as supervisor raters. Unfortunately research has failed to address this gap within the literature, which is why this study looks at performance ratings in team contexts.

Teams, Race, Attributions, and Performance Ratings

The past twenty years have evidenced many distinct changes in the workplace, including a transition from individual to team work, where several employees are designated a single task to complete together, often with a designated leader (Ilgen, 1994). In order to be successful, employees must achieve their goals as members of a cohesive group. But, although teams work as a single unit, they are still composed of individual members, often with varying levels of expertise and ability. Such individual differences make it difficult to judge the performance of another team member, which, if inaccurate, could have deleterious effects on team performance. Team members may not be willing to participate in additional tasks if not satisfied with the performance of one of their members (Phillips, 1997). The question is, how do individuals rate the performance of other team members, specifically the team leader, and is their appraisal equally accurate across team leaders of the same ability?

Little research has been done in order to examine the effects of race on the performance evaluations of team leaders. However, several studies have looked at racial differences in the evaluation of managerial performance (e.g., Davidson, Swigert, and Ruderman, 1998; Mount, Sytsma, Hazucha, and Holt, 1997). Davidson et al. (1998) found that African American managers were rated higher by both White American and African American subordinates. On the other hand, Mount et al. (1997) found that African American subordinates rated managers of the same race higher, while White Americans did not exhibit a racial preference. Unfortunately several issues inhibit the interpretation of their data to the team context. For one thing, managers were involved in their subordinates work to varying degrees depending on the project and the organization. This variance in involvement may have influenced the level of ratings in the studies by both Mount et al. and Davidson et al. Another influence may have been the actual performance level of the managers. Unfortunately the researchers in the aforementioned studies could not control the actual performance level of ratees, which allowed for different interpretations of the same data. Finally, employees are often sensitive to racial issues in the workplace and could have adjusted their written responses accordingly.

Research on gender effects has attempted to more directly examine the rating of team leaders (e.g., Rice et al., 1980; Rosen & Jerdee, 1973), often with conflicting results. Dobbins and Platz (1986) suggested that, in addition to investigating the evaluation of team leaders by gender or any other subgroup, researchers should concentrate on the cognitive process behind the evaluation of different team leaders. In doing so, one general conclusion emerges from the research on the evaluation of team leaders based on gender: behaviors which are consistent with the rater's implicit theories

are more likely to influence evaluations than inconsistent behaviors (e.g., Allen & Ebbesen, 1981; Hastie, 1981; Wyer & Scrull, 1981). In other words, an individual's behavior will be evaluated in accordance with pre-existing subgroup stereotypes, which is the essence of the model proposed in this study.

Although the literature on gender differences mentioned in the previous paragraph concedes that team member behavior is evaluated according to its stereotype consistency, attribution theory has not been introduced as a possible explanation. Conversely, studies that have introduced causal attributions as a possible explanation of team member evaluation have not examined any subgroup differences (e.g., Norvell & Forsyth, 1984; Mitchell & Wood, 1980; Martinko & Gardner, 1987). Attribution researchers have noted "that a comprehensive perspective of the interactive nature of the leader/member attributional processes is needed" (Martinko & Gardner, 1987, p.235). To create such an extensive theory, it is likely that attributions of team leader performance will vary across subgroups. This variance should therefore be included in such models.

Team member ratings of a team leader introduces an important variable, the fundamental attribution error, which could influence the ability to translate attributional information concerning the rating process collected in this study to previous work on subordinate rating (e.g., Kraiger & Ford, 1985). The fundamental attribution error has been extensively documented in the literature (e.g. Ickes & Layden, 1978; Zuckerman, 1979). Basically, individuals will take credit for success and will blame external causes for failure. Although the fundamental attribution error is expected to have an effect, it is not anticipated to change the original model pictured above because it is only likely to impact the strength of the attributional effects. In other words, when seeking blame for

failure, it is expected that individuals will attribute it to internal causes in the case of an African American team leader and to external causes in the case of a White American team leader. On the other hand, when taking credit for the success of the team, it is proposed that individuals will still attribute some of their good fortune to the high performance of their team leader. If the team leader is African American, success is likely to be attributed to external causes, whereas a White American team leader will receive internal attributions as expected. In other words, these effects are predicted to be in the same direction as effects for raters and ratees who are not interdependent.

It is not just a question of who gets the credit for success or failure (i.e., the fundamental attribution error), but also how much they get. In other words, the ratings of the team leader will also be influenced by the amount of blame or praise given to the other team members, not just to the team leader. In other words, raters must apportion blame or praise between all individuals participating in the team. But again, no matter how it is allocated, some amount of blame or praise will necessarily be given to the team leader in the attributional pattern previously described. What it will influence is the level of the direct effect of race on performance ratings. Although African Americans will still be rated lower than White Americans, those individuals who feel the other team members are more responsible than the team leader will give higher ratings to the leader compared to individuals who feel the outcome is entirely the team leader's responsibility.

Hypotheses

The previous model, reproduced below, based on dyadic rating research, is expected to hold in teams where staff members rate their team leader. My hypotheses are

discussed in reference to the links in Figure 8 below. The numbers on the lines of the figure refer to the corresponding hypotheses.

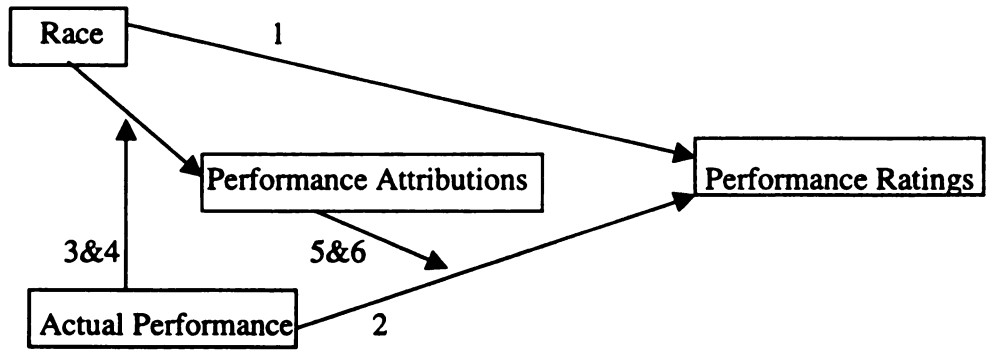

Figure 8. *Model of the relationship between race and performance ratings in a team situation*

Hypothesis 1. There is a relationship between race and performance ratings, such that African American team leaders will receive lower performance ratings than White American team leaders.

Hypothesis 2. There is a relationship between actual performance and performance ratings, such that team leaders in the low performance condition will receive lower performance ratings than team leaders in the high performance condition.

The relationship between race and performance attributions is moderated by actual performance in the following ways:

Hypothesis 3. The relationship between race (African American = 0, White American = 1) and internal attributions will be positive in the high performance condition and negative in the low performance condition (see Figure 9).

Hypothesis 4. The relationship between race and external attributions will be negative in the high performance condition and positive in the low performance condition (see Figure 10).

The relationship between actual performance and performance ratings is moderated by performance attributions in the following way:

Hypothesis 5. The relationship between actual performance and performance ratings will be more positive for high internal attributions than for low internal attributions (see Figure 11).

Hypothesis 6. The relationship between actual performance and performance ratings will be more positive for high internal attributions than for low internal attributions (see Figure 12).

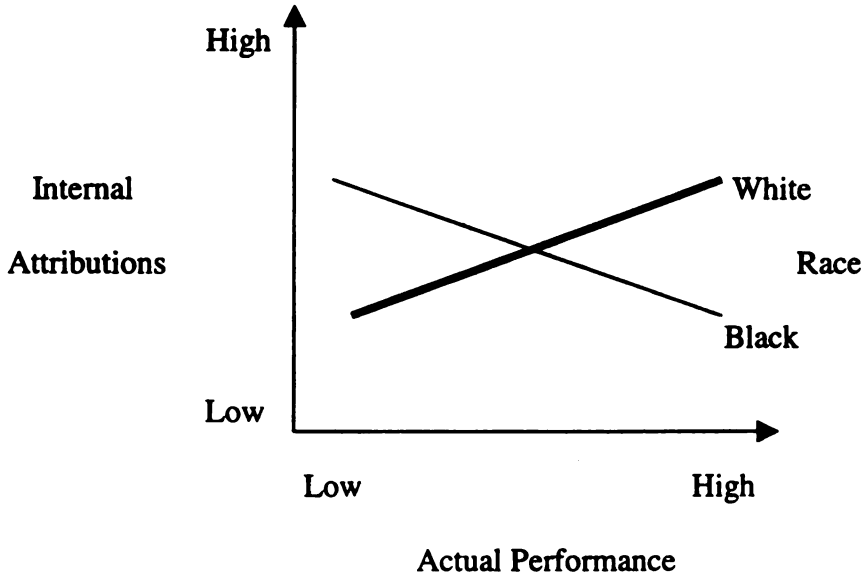

Figure 9. *Hypothesized graph of the interaction between race and actual performance*

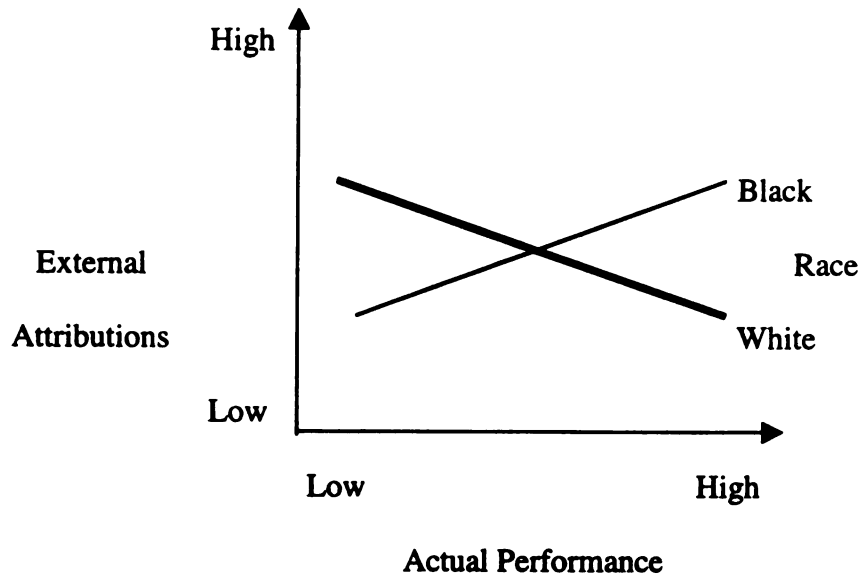

Figure 10. *Second hypothesized graph of the interaction between race and actual performance*

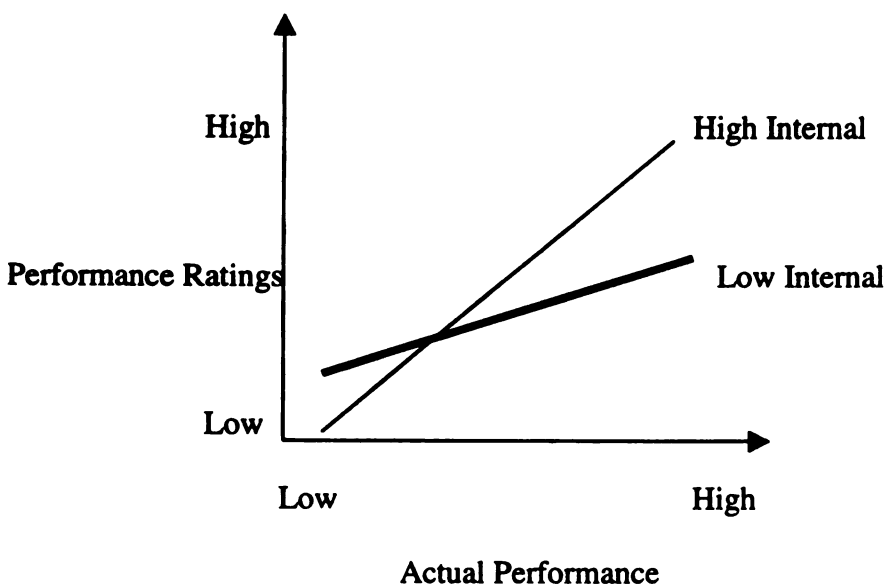

Figure 11. *Hypothesized graph of the interaction between causal attributions and actual performance*

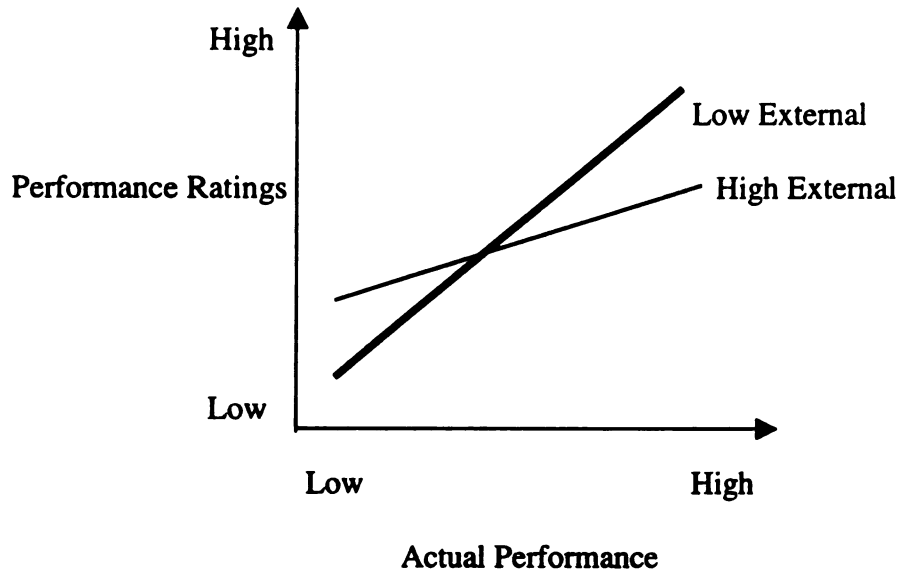

Figure 12. *Second hypothesized graph of the interaction between causal attributions and actual performance*

Method

Participants

One hundred and eighty undergraduates at a large midwestern university participated as team members for three hours in exchange for $20. Each subject was also informed that top scoring team members would have the opportunity to earn bonus money on the basis of their performance. Participants were then placed in a four-person team consisting of three raters and one confederate team leader.

Task

The task the students were presented with was the TIDE^2 decision making task (Team Interactive Decision Exercise for Teams Incorporating Distributed Expertise; Hollenbeck et al., 1995). TIDE^2 is a software program for a decision task that presents

participants with a number of attribute values relevant to either a problem or an object. A judgement regarding that object or problem is then formed based on the values of the attributes. Then decisions are made by the participants regarding the state of that object (e.g., the severity of an injury, the value of a piece of merchandise, etc.). In this study, the software program simulated a naval command-and-control scenario with a team leader and three staff members. The team leader assumed the role of aircraft carrier commanding officer (Alpha), while the staff members assumed the roles of three commanding officers of different air patrol units: an AWACS radar plane (Bravo), an Aegis cruiser (Charlie), and a Coastal Air Defense (Delta).

All four team members sat at separate computer stations, which were networked together. Alpha, either the African American or White American confederate, was seated at a computer station in an adjacent room. This was done for several reasons. First, the confederate team leader entered team decisions from a script, and it was imperative that the staff members were not aware that his behavior was scripted. Second, by having the confederate in another room, his staff members hopefully felt less self-conscious about assigning blame for failure or selfishly taking credit for success. Third, since no verbal communication was allowed during the game, separating the players most likely helped to reduce the amount of talking.

The object of the game was to monitor the airspace surrounding the team. Each team member gathered information regarding particular attributes of an incoming aircraft that entered the airspace monitored by the team. Bravo was in charge of the aircraft's size, speed, radar, and frequency. Charlie was responsible for the aircraft's direction,

speed, angle, and range. And Delta was in charge of the aircraft's radar, altitude, corridor status, and range (see Figure 13).

Bravo	Charlie	Delta
Size Speed Radar Frequency	Direction Speed Angle Range	Radar Altitude Corridor Status Range

Figure 13. *Role requirements*

Bravo, Charlie, and Delta were all instructed that their area of expertise only included their four characteristics, while Alpha's area of expertise included all nine characteristics. In order to make a decision regarding the threat of the aircraft, each staff member had to have information on all four of their characteristics. However, on each target, the staff members could measure four attributes, only one of which was necessary for their rule, while Alpha could measure all nine (See Figure 14). That meant that each staff member had to get three characteristics from his or her teammates. As shown in Figure 15, the staff members had several options to choose from when requesting characteristics from their teammates. However, the option of requesting information from Alpha was discouraged, since he had to measure and evaluate all nine attributes and would not have much time to help his staff members.

After the staff members sent their recommendations to Alpha, he entered the final team decision. All final team decisions and staff member recommendations varied in aggressiveness on a 7-point continuum from Ignore (1) to Defend (7), with defend

indicating the highest level of threat. A clock on the screen counted down the time before the team needed to make a decision. At thirty seconds, the computer began to beep indicating that time was running out. The staff members were told that Alpha could either follow the advice of his team members, or he could make decisions on his own. But instead of combining the three subordinates' recommendations, Alpha made decisions based on prearranged script so as to manipulate the performance level of the leader as seen by the subordinates. Thus, leader behavior was manipulated to create either high or low performance across the experimental sessions.

After the team leader entered the final team decision, a feedback screen appeared on all four computers comparing the team's decision to the correct decision. The team either saw a hit (2 points), a near miss (1 point), a miss (0 points), an incident (-1 points), or a disaster (-2 points). Team members also saw how each of their individual recommendations compared to the correct decision. The team's overall score (an average of all decisions) was given on the bottom of the screen next to the team goal score. The goal score was entered based on the average performance of 90 teams from LePine (1998). All 90 experienced the same target set that was used in this study. The goal score remained constant across both high and low performing teams. Finally, the feedback screen included the team's performance history (i.e., the number of hits, misses, etc.) on the right hand portion of the screen.

Bravo	Charlie	Delta	Alpha
Size Direction Range Corridor Status	Radar Frequency **Angle** Corridor Status	Frequency Speed Direction **Altitude**	Frequency Speed Direction Altitude Radar Angle Corridor Status Size Range

Figure 14. *Measurable attributes*

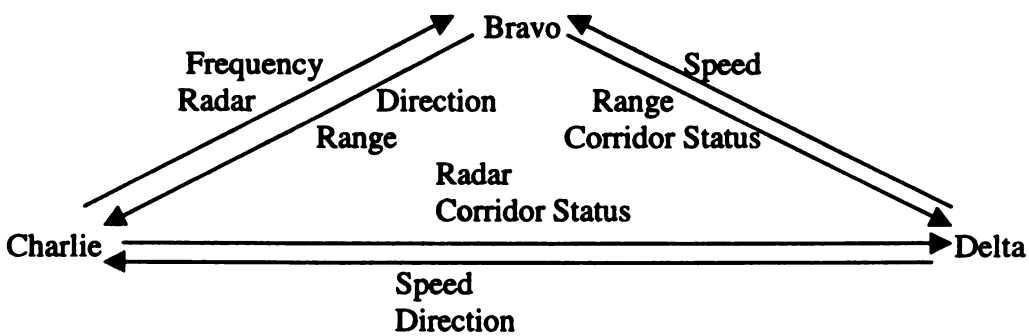

Figure 15. Role structure

Procedure

Participants were recruited via advertisements posted in dorm rooms, classrooms, and the university newspaper. Then they were scheduled to time slots to create teams and randomly selected into conditions. Upon arrival to the laboratory, participants (including the confederate team leader) signed a consent form while they waited for other team members to arrive. After three subjects arrived, they were taken into a separate room along with the confederate team leader. They were then instructed to take a seat at one of the three computer consoles. The confederate team leader walked in last in order to ensure that the subjects occupied all the computers. It was announced that, since he was the last person to enter the room, the confederate would be the team leader and would play the game in another room because verbal communication was not allowed. Once the confederate was escorted out of the room, the participants were welcomed and were given a general task overview booklet and an individual role responsibility sheet. The experimenter then left the room for 8 minutes in order to let them read through the materials. Upon returning, the experimenter answered any questions before beginning the "hands-on" portion of the training where the subjects were presented with five practice trials.

During the first practice trial, the experimenter explained the mechanics involved in gathering and sharing information about target attribute using the pull down menu on their computer screens. The mechanics that were included were (a) measuring the attributes of the aircraft, (b) querying others for attribute values, (c) directly transmitting attribute values to others, and (d) communicating via sentence-long free form text-messages.

The first practice trial also included instruction on individual role responsibilities, which were provided on the role responsibility sheet. The instruction included (a) the specific attributes needed for each role, (b) how to translate raw data on targets into judgments about how threatening the target is based on one attribute, and (c) how to combine all three attributes into a recommendation about a rule. It was also explained that Alpha could measure all nine attributes and could make decisions on each rule without help from his staff members.

After the staff members made their recommendations regarding the first target, the confederate team leader combined their decisions (only for the five practice trials) and made a final team decision. When the feedback screen appeared, the experimenter pointed out (a) how to read the screen (i.e., whether they got a hit, miss, etc.), (b) their overall score, (c) their goal score, (d) how their decisions relate to the actual score, and (e) their performance history.

The practice trials were also used to practice communicating with one another through the text mode and other pre-programmed messages in order to get the information they needed for their role. The experimenter was present to answer any questions that the subjects might have had. All three participants were also reminded that any communication must occur over the computer. After the last practice trial, the experimenter wished the team "good luck" and started the game clock. An "aircraft" soon appeared in their airspace, and the team began to gather and process the relevant information.

Measures and Manipulations

Race. The team leader's race was manipulated by randomly assigning either the African American or the White American confederate team leader to either the high or low performance condition. Both team leaders were male and similar in age, height, and weight.

Actual Performance. The actual performance of the team was manipulated. The goal score for teams in both the high and low performance conditions was equal to the average score of the 90 teams from LePine (1998). Individuals in the high performance condition scored two standard deviations higher than the average, while individuals in the low performance condition scored two standard deviations lower than the average. The performance levels were then piloted using 18 subjects so as to have levels that were believable but also clearly high or low.

Often there are real differences in performance between African American and White American employees. For example, African Americans tend to score around one standard deviation lower than White Americans on standard cognitive ability tests (Jensen, 1980). This issue will not be addressed by this study, since I am interested in racial bias in ratings and not actual performance differences.

Causal Attributions. In the literature, researchers have used a variety of attribution measures. Unfortunately many of them have exhibited poor reliability and construct validity. Based upon Weiner's (1985) model of performance attributions, an attribution measure was constructed with items specifically referenced to the $TIDE^2$ task features experienced by the participants. Using an 8-item scale, internal attributions were assessed by asking participants whether their team leader's behavior was due to his effort or his

intelligence in the high performance case, or his lack of effort or lack of intelligence in the low performance case. External attributions were assessed in the high performance condition by asking subjects whether the team leader's behavior was caused by luck or the ease of the game. In the low performance condition they were assessed by asking whether the team leader's behavior was due to bad luck or the difficulty of the game. This left four items which measured internal attributions and four items which measured external attributions (see Appendix A). Responses were on a five point, Likert-type scale. Intercorrelations between the eight items are presented in Table 1. Since the internal items correlated more highly with themselves rather than with the external items, two scales of four items each were constructed by summing the four z-scores. Internal consistency for the internal scale, measured using coefficient alpha, was .81. Alpha for the external scale was .67. An exploratory factor analysis on the eight items revealed two factors with eigenvalues greater than 1.0. The four internal items loaded highest on the first factor that explained 35.5% of the variance. The four external items loaded highest on the second factor and explained 23.1% of the variance (see Table 2). A confirmatory factor analysis was also completed. The internal and external latent variables had four indicators each, as shown in Figure 16. The model was tested using AMOS. The fit indices were as follows: GFI=.91; AGFI=.84; RMSEA=.118. The results indicated a fair fit of the data to the hypothesized model. Finally, the two scales evidenced good divergent and convergent validity. The external scale correlated -.18 ($p<.05$) with the internal scale and -.18 ($p<.05$) with three internality items from Russell's (1982) attribution scale. A high score on Russell's three-item scale would indicate that the participant believes that the team leader's behavior was due to internal factors. This was

Table 1

Intercorrelation Between Internal and External Attribution Items

Variable	Mean	SD	i1	i2	i3	i4	e1	e2	e3	e4
1. i1	3.63	.98	--							
2. i2	3.28	.97	.51**	--						
3. i3	3.58	.99	.57**	.41**	--					
4. i4	3.61	.94	.53**	.45**	.60**	--				
5. e1	3.26	1.09	-.06	.02	-.05	-.15*	--			
6. e2	2.68	1.12	.04	.24**	-.07	-.09	.46**	--		
7. e3	3.04	1.09	-.12	-.13	-.10	-.26**	.21**	.15*	--	
8. e4	3.06	1.12	-.22**	-.15*	-.23**	-.31**	.38**	.30**	.51**	--

Note: N=180 i=internal. e=external. * $p < .05$. ** $p < .01$.

Table 2

Factor Categories Derived from Factor Analysis of 8 Causal Attribution Items

Factor	Item	Factor Loading
Factor 1: Internal Attributions		
	i1	.812
	i2	.767
	i3	.780
	i4	.777
Factor 2: External Attributions		
	e1	.749
	e2	.734
	e3	.583
	e4	.728

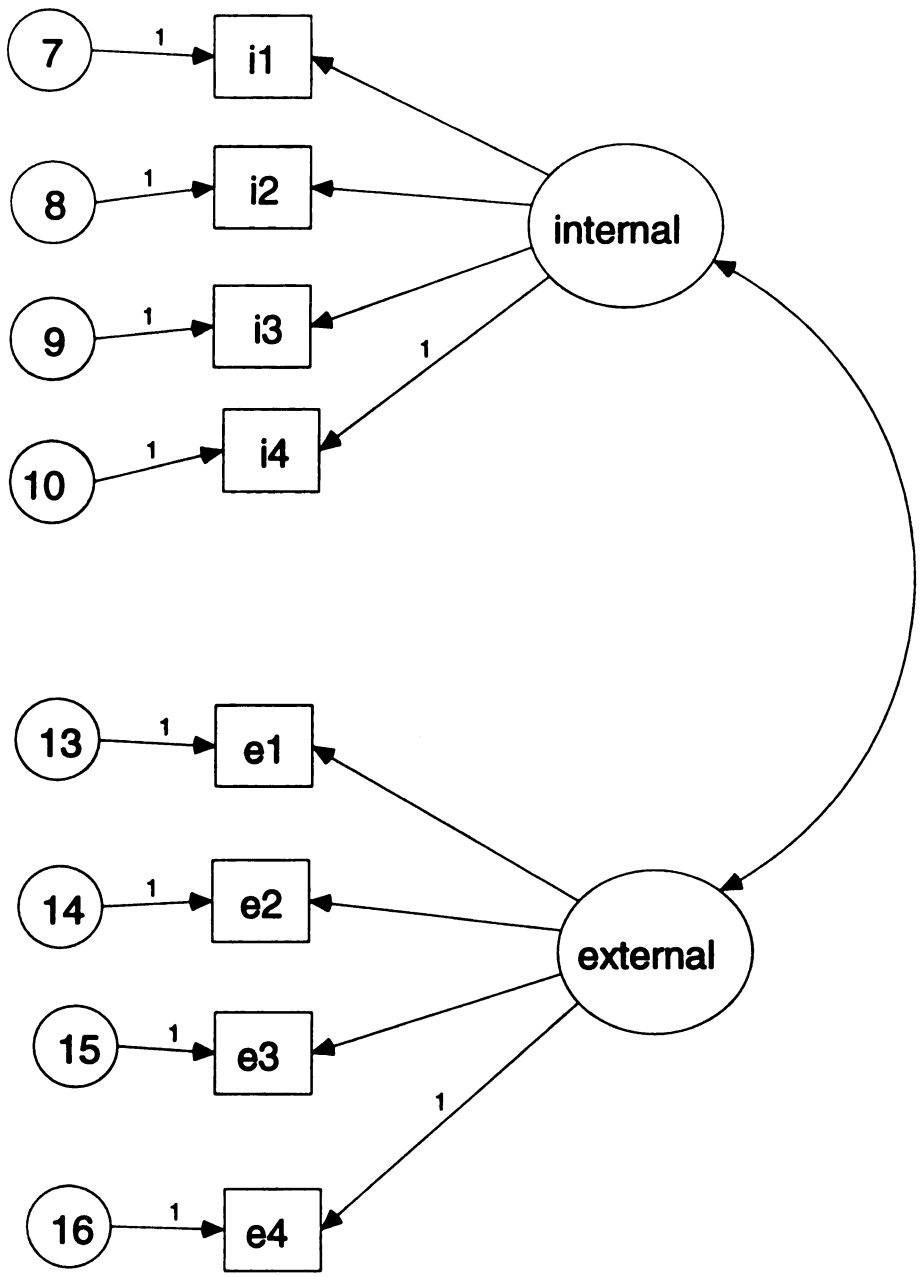

Figure 16. *Confirmatory factor analysis of 8 attribution items*

supported by the fact that the internal scale constructed in this study correlated .29 ($p<.01$) with Russell's attribution scale. Overall, the data support the use of a separate internal and external scales.

Performance Ratings. A 19-item performance measure was constructed in order to evaluate the team leader's performance on TIDE[2]. It included items such as, "Alpha was a good team leader" and "Our team leader performed very well on this task." Responses were on a five point, Likert-type scale (see Appendix B). Coefficient alpha for the scale was .95. Two additional items asked the subordinates, on a scale of 1 to 100, how good their team leader was (see Appendix C). Since the two items scaled from 1 to 100 correlated highly with each other and the 19 item scale (see Table 3), these two plus the sum of the 19 were transformed into z-scores and combined into one leader performance rating scale.

Table 3

Intercorrelation Between 19 Item Performance Rating Scale and 2 Rating Items Scored 1 to 100

Variable	Mean	SD	i1	i2	i3
1. pr19	58.53	15.32	--		
2. pr1	63.43	24.92	.846**	--	
8. pr2	63.89	25.93	.815**	.902**	--

Note: N=180. pr19=19 item performance rating scale. pr1=first rating item scaled 1 to 100.
pr2=second rating item scaled 1 to 100. * $p < .05$. ** $p < .01$.

Results

Table 4 provides the means, standard deviations, and intercorrelations among all the variables used in testing the hypotheses. Table 4 also includes two variables,

subordinate race and subordinate performance, which were not part of the hypotheses but were used in supplemental analyses.

The analyses, as described in detail below, were completed in four distinct steps. First, data gathered as manipulation checks are presented for the team leader's race and the team's actual performance. Second, the analyses used to test each of the hypotheses presented earlier are provided. Third, subordinate race and subordinate performance are added to the analyses to test for possible confounds. Fourth, the results from additional analyses that were performed in order to test the fit of alternative models are presented.

Manipulation Checks

Race. The team leader's race was manipulated depending on whether the subjects were randomly selected into the African American or White American team leader condition. Upon completion of the experiment, subjects were given a two-item questionnaire asking whether their team leader was male or female and African American or White American. Ninety-nine percent of the subjects correctly identified their team leader as an African American or White American male, indicating that the manipulation was effective. Those subjects who incorrectly identified the team leader were included in the analyses to be conservative.

Team Performance. The team's performance was also manipulated depending on whether the team was randomly selected into the low performance or high performance condition. At the end of the experiment, subjects completed a one-item questionnaire asking whether the team performed well or not well. Ninety-five percent of the subjects correctly identified their performance as high or low, indicating that the performance

Table 4

Means, Standard Deviations, and Intercorrelations Among All Analyzed Variables

Variable	Mean	SD	1	2	3	4	5	6	7
1. Actual Team Performance	.50	.50	--						
2. Leader Race	.50	.50	.00	--					
3. Internal Attributions	0.0	3.18	.49**	-.18*	--				
4. External Attributions	0.0	2.83	-.57**	-.01	-.18*	--			
5. Performance Ratings	0.0	2.85	.76**	.00	.15*	-.57**	--		
6. Subordinate Race	1.1	.28	.02	.02	-.04	.04	-.05	--	
7. Subordinate Performance	66.20	12.82	.14	.08	.14	-.07	.06	-.15*	--

Note: N=180. * $p < .05$. ** $p < .01$. Leader Race was coded 0 for the African American leader and 1 for the White American leader. Actual team performance was coded 0 for low performance and 1 for high performance.

manipulation was effective. Again the conservative position was taken by including all subjects in the analyses.

Tests of Hypotheses

Hypothesis 1. The first hypothesis predicted that race would be related to performance ratings, such that African Americans would receive lower ratings than White Americans. Examining the correlation matrix presented in Table 4 shows that race was uncorrelated ($r = .00$, p=ns) with performance ratings. Hypothesis 1 was not supported.

Hypothesis 2. The second hypothesis asserted that leader performance ratings would be related to the team's actual performance, such that team leaders in the low performance condition would be rated lower than team leaders in the high performance condition. Table 4 shows a correlation of .76 ($p<.01$) between actual performance and leader performance ratings. Clearly subordinates rated team leaders in line with the performance information they received on them. Hypothesis 2 was strongly supported.

Hypothesis 3. The third hypothesis predicted that the relationship between race and internal attributions would be moderated by the team's actual performance. To test this hypothesis, a two-step hierarchical regression was performed entering race and actual performance in the first step, and their interaction in the second. Actual performance did have a direct effect on internal attributions ($ß=.56$, $p<.01$), such that subordinates in the high performance condition attributed more of their team leader's performance to internal attributions. However, team leader race was not directly related to internal attributions ($ß=-.10$, $p<.26$) and the interaction between team leader race and actual performance explained an insignificant portion of the variance in internal attributions ($\Delta R^2=.01$, p=ns).

Thus, Hypothesis 3 was not supported. The results of the regression analysis are presented in Table 5.

Table 5

Result of Regressing Internal Attributions on Leader Race, Actual Team Performance, and Their Interaction.

Variable	β	R^2	ΔR^2
Step 1			
Leader Race	-.10		
Actual Team Performance	.56**	.27**	
Step 2			
Leader Race X Actual Team Performance	-.13	.27**	.01

Note: N=180. β is the standardized regression coefficient. Increments for variables entered at the R^2 significance levels are based on F tests for that step. * $p < .05$. ** $p < .01$.

Hypothesis 4. Hypothesis 4 proposed that the relationship between race and external attributions would be negative in the high performance condition and positive in the low performance condition. Once again, a two-step hierarchical regression was performed entering race and actual performance in the first step, and their interaction in the second. Actual performance did have a direct effect on external attributions (ß= -.73, p<.01). As performance increased, subordinates attributed less of their team leader's behavior to external attributions. In addition, race was directly related to external attributions (ß=-.17, p<.05). Specifically, subordinates were more likely to attribute the leader's performance to external causes if the team leader was African American. Finally, the interaction between race and actual performance explained a significant portion of the variance in external attributions (ΔR^2=.03, p<.01). The results of the regression analysis are presented in Table 6.

Table 6

Result of Regressing External Attributions on Leader Race, Actual Team Performance, and Their Interaction.

Variable	β	R^2	ΔR^2
Step 1			
Leader Race	-.17*		
Actual Team Performance	-.73**	.32**	
Step 2			
Leader Race X Actual Team Performance	.28	.35**	.03**

Note: N=180. β is the standardized regression coefficient. Increments for variables entered at the R^2 significance levels are based on F tests for that step. * $p < .05$. ** $p < .01$.

However, although the interaction was significant, Hypothesis 4 was not supported. As illustrated in Figure 17, as performance increases, the slope of the African Americans' regression line was significantly more negative than the line for White Americans. In other words, as performance increased, subordinates were more likely to attribute their team leader's performance to external factors if he were White American.

Hypothesis 5. Although it was expected (and observed) that actual team performance would have a major impact on leader ratings, Hypothesis 5 predicted that the size of the impact would vary depending upon attributions for leader performance. Specifically this hypothesis predicted that the effects of actual performance on performance ratings would be moderated by internal attributions. Again, a two-step hierarchical regression was performed entering internal attributions and actual performance in the first step, and their interaction in the second. As previously noted, actual performance did have a direct effect on performance ratings (ß=.84, p<.01).

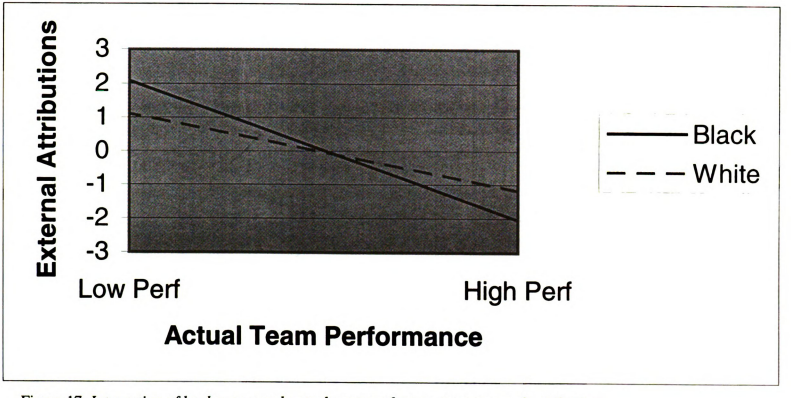

Figure 17. *Interaction of leader race and actual team performance on external attributions*

Furthermore, internal attributions were directly related to performance ratings (ß=-.53, p<.01), such that, as actual performance increased, subordinates were less likely to highly rate their team leader if they attributed his behavior to internal causes. Finally, the interaction between internal attributions and actual performance explained a significant portion of the variance in performance ratings (ΔR^2=.09, p<.01). The results of the regression analysis are presented in Table 7.

Table 7

Result of Regressing Leader Performance Ratings on Actual Team Performance, Internal Attributions, and Their Interaction.

Variable	β	R^2	ΔR^2
Step 1			
Actual Team Performance	.84**		
Internal Attributions	-.53**	.63**	
Step 2			
Internal Attributions X Actual Team Performance	.41	.72**	.09**

Note: N=180. β is the standardized regression coefficient. Increments for variables entered at the R^2 significance levels are based on F tests for that step. * p < .05. ** p < .01.

Figure 18 shows that, as performance increases, the slope of the regression line for high internal attributions was significantly more positive. But for low internal attributions, performance ratings were unrelated to actual performance. Thus, Hypothesis 5 was supported.

Hypothesis 6. Hypothesis 6 suggested that the effects of actual performance on performance ratings would be moderated by external attributions. As in the previous hypotheses, a two-step hierarchical regression was performed entering external

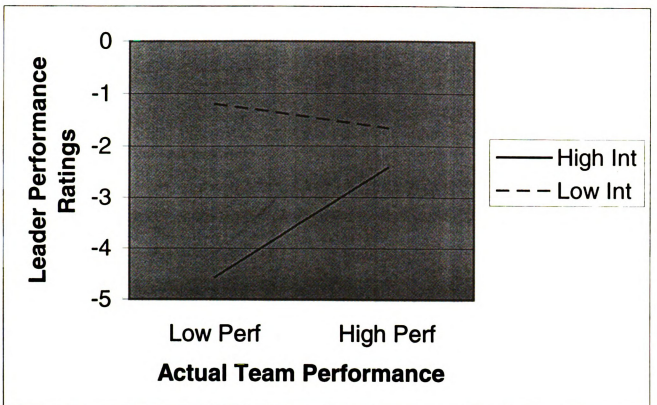

Figure 18. *Interaction of internal attributions and actual performance on leader performance ratings*

attributions and actual performance in the first step, and their interaction in the second. Actual performance again had a direct effect on performance ratings (ß=.64, p<.01). In addition, external attributions had a direct effect on performance ratings (ß=-.27, p<.01), such that subordinates were less likely to rate their team leader higher as performance increased if they attributed his behavior to external causes. However, the interaction between external attributions and actual performance did not explain a significant portion of the variance in performance ratings (ΔR^2=.00, ns). The results of the regression analysis are presented in Table 8. Thus, Hypothesis 6 was not supported.

Table 8

Result of Regressing Leader Performance Ratings on Actual Team Performance, External Attributions, and Their Interaction.

Variable	β	R^2	ΔR^2
Step 1			
Actual Team Performance	.64**		
External Attributions	-.27**	.60**	
Step 2			
External Attributions X Actual Team Performance	.08	.60**	.00

Note: N=180. β is the standardized regression coefficient. Increments for variables entered at the R^2 significance levels are based on F tests for that step. * p < .05. ** p < .01.

Summary of Hypotheses

In conclusion, only two of six hypotheses were supported by the data. One possible explanation could have been the influence of extraneous variables. The subordinate's race and the subordinate's level of performance could have had an effect on the results presented above. In particular, the subordinate's race could have effected the relationship between leader race and performance ratings. It is reasonable to suggest that

African Americans rate team leaders of their own race differently than if they were part of the majority. It is also reasonable to suggest that the effect of race on causal attributions will depend not only on the team's actual performance level, but also on the ratee's race in relation to that of the rater. African Americans may attribute behavior to different causal loci than White Americans when the team leader is a minority.

In addition, since both the rater and the ratee were doing similar tasks, it is possible that the raters' own performance may have influenced their responses on the attribution and performance rating measures. For example, someone who performed very poorly on the game could have given the team leader positive ratings even if they were in the low performance condition. In addition, someone who had difficulty with TIDE[2] might be more likely to attribute low performance to external causes than someone who did very well at the game. As a result, the subordinate's performance level and the subordinate's race were entered into the supplemental analyses carried out on the data.

Supplemental Analyses

Subordinate Race. It was thought that subordinate race might influence leader performance ratings. However, due to the small sample of African American subordinates (n=15), the data lacked sufficient power to test this well. In fact, race did not relate to performance ratings (r=-.05, ns). It also did not influence the relationship between the race of the leader and the leader's performance ratings (ß=-.35, ns). Finally, the subordinate's race did not interact with the leader's race and actual performance to effect either external (ß=.35, ns) or internal (ß=.23, ns) attributions.

Subordinate Performance. The second variable, the subordinate's level of performance, was explored as it related to performance ratings. Again, there was no

evidence of a direct effect (r=.07, ns). And the subordinate's performance did not influence the relationship between the team's actual performance and performance ratings (ΔR^2=.003, ns). Lastly, the subordinate's level of performance did not interact with the leader's race and actual performance to impact either external (ß=.04, ns) or internal (ß=-.07, ns) attributions.

To sum up, the supplemental analyses of the subordinate's race and performance level did not help to clarify the previous results. However, the analyses were all based on the model presented in the introduction (see Figure 7). It could be that an alternative model explains the data much better. As described in the introduction, many of the studies examining the link between race and performance ratings have found conflicting results (e.g., Kraiger & Ford, 1985; Bass & Turner, 1973). However, none of the aforementioned researchers attempted to collect laboratory data where the performance of the ratee could be manipulated. Therefore, their results are subject to actual differences in performance, which could effect the relationship of race and performance ratings (e.g., Kraiger & Ford, 1985).

Post Hoc Analyses

To investigate whether the relationship between leader race and performance ratings depended on the team's actual performance, a two-step hierarchical regression was set up entering leader race and actual performance in the first step, and their interaction in the second. As stated in earlier analyses, actual performance effected performance ratings (ß=.86, p<.01) and leader race did not effect performance ratings (ß=.10, ns). However, the interaction between leader race and actual performance did

explain a significant portion of variance in performance ratings (ΔR^2=.01, p<.05). The results of the regression analyses are presented in Table 9.

Table 9

Result of Regressing Leader Performance Ratings on Leader Race, Actual Team Performance, and Their Interaction.

Variable	β	R^2	ΔR^2
Step 1			
Actual Team Performance	.86**		
Leader Race	.10	.57**	
Step 2			
Leader Race X Actual Team Performance	-.18*	.58**	.01*

Note: N=180. β is the standardized regression coefficient. Increments for variables entered at the R^2 significance levels are based on F tests for that step. * p < .05. ** p < .01.

However, the interaction was not as expected. As shown in Figure 19, African American team leaders' ratings were higher than White American team leaders' ratings under high performance. In the low performance condition, African Americans were rated slightly lower than White Americans were. In the high performance condition, on the other hand, African Americans were rated much higher than White Americans were.

The next step in the post hoc analyses involved the addition of causal attributions to the model. It is possible that African Americans were rated higher in the high performance condition because their behavior was attributed internally, and this in turn could have contributed to their level of performance ratings. So internal attributions as mediators of the relationship between race and performance ratings were explored in the high performance condition. In order to test for mediation, it is necessary to demonstrate that (a) both the independent (leader race) and the mediating (internal attributions)

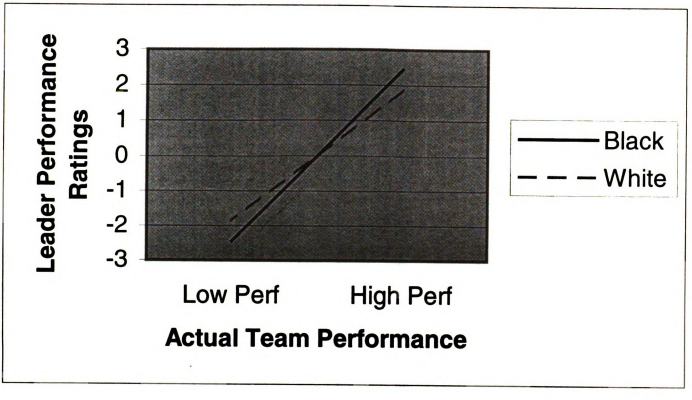

Figure 19. *Interaction of leader race and actual team performance on leader performance ratings*

variables relate to the dependent variable (performance ratings), (b) the independent variable relates to the mediating variables, (c) the relationship between the independent variable and the dependent variable becomes negligible or is reduced significantly when controlling for the mediating variables, and (d) the relationship between the mediator variables and the dependent variable is still significant when controlling for the independent variable (Baron & Kenny, 1986; James & Brett, 1984).

Both race ($r=-.23$, $p<.05$) and internal attributions ($r=.32$, $p<01$) were related to performance ratings, which satisfied the first condition of mediation (see Table 10). Race was also related to internal attributions ($r=-.36$, $p<.01$), which satisfied the second condition (see Table 10).

Table 10

Means, Standard Deviations, and Intercorrelations Among All Variables Analyzed Post Hoc in the High Performance Condition

Variable	Mean	SD	1	2	3	4
1. Leader Race	.50	.50	--			
2. Internal Attributions	1.54	2.30	-.36**	--		
3. External Attributions	-1.60	2.21	.19	-.07	--	
4. Leader Performance Ratings	2.15	1.22	-.23*	.32**	-.26*	--

Note: N=90. * $p < .05$. ** $p < .01$.

Hierarchical regression was employed to test the remaining conditions. As shown in Table 11, when internal attributions were entered in the first step, leader race was no longer significantly related to performance ratings, which passes the third mediation hurdle. Finally, the fourth condition was also satisfied because the relationship between internal attributions and performance ratings remained significant even when controlling for leader race (see Table 12). So, in the high performance condition, internal attributions

Table 11

Result of Regressing Leader Performance Ratings on Leader Race while Controlling for Internal Attributions in the High Performance Condition

Variable	β	R^2	ΔR^2
Step 1			
Internal Attributions	.27*	.10**	
Step 2			
Leader Race	-.14	.12**	.02

Note: N=90. β is the standardized regression coefficient. Increments for variables entered at the R^2 significance levels are based on F tests for that step. * $p < .05$. ** $p < .01$.

Table 12

Result of Regressing Leader Performance Ratings on Internal Attributions while Controlling for Leader Race in the High Performance Condition

Variable	β	R^2	ΔR^2
Step 1			
Leader Race	-.14	.05*	
Step 2			
Internal Attributions	.27*	.12**	.06*

Note: N=90. β is the standardized regression coefficient. Increments for variables entered at the R^2 significance levels are based on F tests for that step. * $p < .05$. ** $p < .01$.

mediated the relationship between leader race and performance ratings as initially predicted across all levels of performance. Subordinates were more likely to attribute the performance of the African American team leader to internal attributes and this in turn increased their performance ratings.

It is also possible that White Americans could have been rated lower in the high performance condition because their behavior was attributed externally, and this in turn contributed to their level of performance ratings. So it was further hypothesized that external attributions mediated the relationship between race and performance ratings in the high performance condition. As previously noted, leader race was related to performance ratings. And since external attributions were also related to the dependent variable ($r=-.259$, $p<.05$), the first condition was satisfied (see Table 10). The second condition required external attributions to correlate with the leader's race ($r=.193$, $p<.07$), which passed (see Table 10). Hierarchical regression was again used for the remaining steps. When external attributions were entered in the first step, leader race was no longer significantly related to performance ratings, which satisfied the third condition (see Table 13). The data also passed the fourth and final condition, since the relationship between external attributions and performance ratings remained significant even when controlling for leader race (see Table 14). Therefore, in the high performance condition, external attributions mediated the relationship between leader race and performance ratings. Subordinates were more likely to attribute the performance of the White American team leader to external attributes and this in turn decreased their performance ratings.

Table 13

Result of Regressing Leader Performance Ratings on Leader Race while Controlling for External Attributions in the High Performance Condition

Variable	β	R^2	ΔR^2
Step 1			
External Attributions	-.22	.07*	
Step 2			
Leader Race	-.19	.10**	.04

Note: N=90. β is the standardized regression coefficient. Increments for variables entered at the R^2 significance levels are based on F tests for that step. * $p < .05$. ** $p < .01$.

Table 14

Result of Regressing Leader Performance Ratings on External Attributions while Controlling for Leader Race in the High Performance Condition

Variable	β	R^2	ΔR^2
Step 1			
Leader Race	-.19	.05*	
Step 2			
External Attributions	.22*	.12**	.05*

Note: N=90. β is the standardized regression coefficient. Increments for variables entered at the R^2 significance levels are based on F tests for that step. * $p < .05$. ** $p < .01$.

In the low performance condition, the data did not support the mediation predictions. Since leader race was not significantly related to performance ratings (r=.126, ns), the first condition of mediation was not satisfied. As a result, further analyses on subsequent steps would have been meaningless.

Discussion

The main goal of this study was to examine the relationship between race and performance ratings from a different perspective than has been the norm in the literature. Most research has focused on whether African American and White American subordinates are evaluated equally by their superiors (e.g., Kraiger & Ford, 1985; Sackett & DuBois, 1991). Overall, it appears that African American subordinates are rated lower than their White counterparts (Sackett & DuBois, 1991).

However, within the organization there are alternative sources of feedback that are now being used instead of, or supplemental to, the standard performance appraisal by a superior (e.g., Miles & Snow, 1995). Researchers have found that subordinates and peers also have critical information to offer an employee regarding his or her performance (e.g., Borman, 1974).

In addition, there has also been a transition within the workplace from individual to team work, where a task is completed by a group of individuals, often with a designated team leader (Ilgen, 1994). In order to evaluate the effectiveness of a team leader, it is essential that as much information as possible is collected regarding his or her effectiveness, including performance ratings from the members of the team.

Finally, most of the literature is based on data for which there is little or no ability to assess the impact of the performer's actual performance on ratings. Since this study

was able to control the level of performance of the team leader, it offered an advantage by examining ratings given to a very poor versus very successful team leader. Researchers have noted that, in order to better estimate the population correlation between race and performance ratings, it is necessary to separate out the ratee's actual level of performance (e.g., Pulakos et al., 1989).

Since the literature offered little empirical work regarding the relationship between team leader race and subordinate performance ratings, it was assumed that race would effect ratings in a similar fashion to Sackett and DuBois (1991). It was expected that African American team leaders would get rated lower than their White counterparts. However, despite the control offered by the lab, this study found team leader race was unrelated to performance ratings across all performance conditions (r=.00, p=ns). Clearly the results described by Sackett & DuBois (1991) did not translate to this team environment where subordinates evaluated the performance of their team leader.

The second goal of this study was to investigate the role of performance attributions in possible cognitive biases present in the evaluation of African American and White American team leaders. It has been suggested that attribution theory (Weiner, 1985) could help explain why African American subordinates are rated lower than White American subordinates when their performance is evaluated by a superior within the workplace (Ilgen & Youtz, 1986). Therefore it was hypothesized that attribution theory may also influence the relationship between race and performance ratings when the ratee is the superior and not the subordinate. It was expected that African Americans would be rated lower in the low performance condition because their behavior would be attributed internally, while they would be rated lower in the high performance condition because

their behavior would be attributed externally. However, no relationship was found between team leader race and the evaluation of his performance. This negated any further investigation of the relationship using attribution theory. Our model did not fit the data.

There are several possible explanations why the model presented in the introduction was not supported. First, it could be that the effects of race on performance ratings wane when raters and ratees work collectively on a designated task. By interacting with one another for an extended period of time, the leader's race could have lost its impact when each team member rated their leader's performance. But if the team members did not evaluate their leader on the basis of his race, then their behavior should be attributed to similar causes. For example, poor performance should be attributed to external factors for both African American and White American team leaders. However, the data regarding external attributions did not support such an assertion. In the low performance condition, African Americans were more likely to have their behavior attributed to external causes. And, in the high performance condition, White Americans were more likely to have their behavior attributed to external causes.

The data regarding internal attributions were a little more encouraging, since race did not have an impact on whether the leader's behavior was attributed internally or externally. And when the team leader's poor performance was highly attributed to internal causes, he was rated much lower. Although that difference dissipated when the team was successful, the team leader still received lower ratings when his behavior was highly attributed to internal causes.

In conclusion, causal attributions offered little explanation as to the reason why African Americans and White Americans were rated similarly. African American and

White American team leaders did not always receive similar causal attributions for their performance and those attributions did not effect performance ratings as expected.

A second explanation of why the hypothesized model was not supported involves the level of analytical power available in this study. It could be that a sample size of 180 subjects was simply too small to obtain a significant relationship between race and performance ratings. Much of the latest research on race and performance ratings has relied on the technique of meta-analysis, which combines the data from a variety of studies completed in the field and in the laboratory in order to come up with a better estimation of the population correlation (Hunter, Schmidt, & Jackson, 1982). By doing so, researchers have been able to analyze the relationship between race and performance ratings using thousands of subjects (e.g., Sackett & DuBois, 1991; Kraiger & Ford, 1985). This study, on the other hand, dealt with only a small sample of undergraduate students and that could have affected my ability to find small effect sizes. However, in this study the relationship between race and performance ratings was almost nil, which means that a substantial raise in power would probably not have helped the correlation reach significance.

Since power was not likely an issue, a third possible explanation was the existence of a moderator. In this study, one potential moderator of the relationship between race and performance ratings was the actual performance of the team. As stated earlier, laboratory research offers the advantage of increased control. By using field data, researchers lose the ability to control and manipulate variables. Much of the literature on the relationship between race and performance ratings has dealt with samples gathered from the field. For example, in Kraiger and Ford's (1985) meta-analysis, they note that

their results do "not preclude the possibility that actual performance differences between races existed" (pp. 62). For many of the early studies, it was impossible to determine whether African Americans were rated differently due to bias or due to the fact that they actually performed their job less effectively. As a result, some of the studies found a no relationship between race and performance ratings, while others found African Americans and White Americans rated individuals of the same race higher (e.g., Hamner et al., 1974; Bass & Turner, 1973). Pulakos et al. (1989) proposed that any bias in ratings given to African American and White American employees had to be separated from their actual level of performance. In fact, researchers have shown that the relationship between race and performance ratings depends on the actual performance of the ratee (Bigoness, 1976; Rotter & Rotter, 1969). Following along those same lines, this study proposed post hoc that the absence of a statistically significant relationship between race and performance ratings may have been due to the moderating effect of ratee performance.

A new model was tested investigating whether the actual performance of the team leader moderated the relationship between leader race and performance ratings. As can be seen in Figure 19, there was a significant interaction between the leader's race and his actual performance. The African American team leader was rated higher than the White American team leader in the high performance condition and lower in the low performance condition. However, in the low performance condition, the relationship between leader race and performance ratings did not reach significance (r=.13, p=ns). Based on past research (e.g., Sackett & DuBois, 1991), it was expected that African Americans would be rated lower due to the stereotypes overarching the African American

community (e.g., Smith, 1990). Specifically, stereotypes for African Americans are of less ambition and lower intelligence than White Americans. This view of African Americans creates expectations concerning their behavior, and if an African American performs poorly on a certain task, the behavior is consistent with the stereotype. Yet in this case both team leaders were rated almost equally, with the African American team leader receiving slightly lower ratings.

One possible explanation for this could be a lack of power. By splitting the sample into only the low performance condition, the power to find a significant relationship dropped by fifty percent. Perhaps if the number of subjects doubled, the relationship would have been significant.

Another explanation might be the students' inclination to give socially desirable responses (Jones & Signall, 1971), thereby ameliorating the amount of blame given to the African American team leaders. Most people are aware of the stereotypes assigned to African Americans, and, in an effort to stay away from such negative responses, the staff members could have unconsciously or consciously raised their ratings of the African American team leader. And since the consequences for poor performance were relatively minor (i.e., everyone got paid $20 no matter how well they performed), it may have been easier to tone down the ratings given to the African American team leader.

The reverse may also have been true. The stereotypes of high intelligence and effort given to White culture could have severely impacted the performance ratings given to the White team leader after he mishandled target after target on TIDE^2. It was proposed that White American team leaders would receive the benefit of the doubt in the low performance condition, and their performance would be attributed to bad luck or the

difficulty of the game. In that way, their performance ratings would not suffer as much as the African American team leaders. But in this case, the staff members may have found the White American team leader's poor performance on such an easy task so salient that the only explanation was complete incompetence.

Although African Americans were only rated slightly lower than Whites in the low performance condition, race was significantly correlated with performance ratings ($r=-.233$, $p<.05$) in the high performance condition. African Americans were rated higher than White Americans when they led their team to success. Originally it was proposed that the stereotypes surrounding African Americans would prevent them from receiving any credit for success while playing TIDE^2. The African American team leaders' success would simply be blamed on luck or the ease of the game. As a result, by trivializing the excellent performance of the African American team leader, the team members would have enough cause to effectively lower their performance ratings. However, in this study, the performance of the African American team leader was not trivialized, but was applauded.

Nevertheless, it is still possible that the team members did not expect their African American team leader to do so well on TIDE^2, since it contradicted their stereotypes of African American culture. Instead of trying to suppress conflicting information by attributing the behavior to external causes, the team members may have resolved the conflict by viewing his performance as an exception to the rule. As a result, the subordinates might have attributed the success of the African American team leader to internal factors such as intelligence and effort, and that in turn bolstered their performance ratings. In fact, the results of the mediation analyses supported this

assertion. Internal attributions mediated the relationship between race and performance ratings, such that race was negatively correlated with internal attributions and positively correlated with performance ratings. Therefore, African Americans received internal attributions for their behavior in the high performance condition, and that resulted in higher performance ratings.

If the subordinates were expecting their African American team leader to fail to a greater extent than those with a White team leader, then it is also possible that they were expecting the White American team leader to succeed. If so, in the high performance condition, the White American team leader does just as expected by hitting target after target on TIDE^2. But fulfilling expectations does not usually reap the same reward as surpassing expectations. Maybe the subordinates felt that their White American team leader did not accomplish anything remarkable during the game and thus did not deserve to receive all the credit for his performance. The results do, in fact, support this contention, since the successful performance of the White team leader was more likely to be attributed to external causes such as luck or the difficulty of the game. And the directions of the correlations imply that that may be why the White American team leader was rated lower in the high performance condition.

Theoretical Implications

The results of this study have both theoretical and practical implications. Theoretically, it is clear that male ratees' level of performance influences the relationship between their race and their performance ratings, as other researchers have suggested (e.g., Pulakos, 1989). This helps to explain why some researchers investigating the ratings of managers by their subordinates find conflicting results. For example, Davidson

et al. (1998) found that African American managers were rated higher than White American managers, while Mount et al. (1997) found that White subordinates rated African American and White American managers similarly. It could have been the actual performance of the managers that influenced the differential effects.

But even if the actual performance of the ratee moderates the relationship between race and performance ratings, it is clear that it depends on the rating situation. For example, several studies have shown that, when performance is low, African Americans are rated higher than White Americans are, and, when performance is high, there is no relationship between race and performance ratings (Bigoness, 1976; Rotter & Rotter, 1969). In addition, other studies controlling for the actual performance of the ratee have found that African Americans are actually rated lower than White Americans (e.g., Sackett & DuBois, 1991). No studies have shown that African Americans are rated higher than White Americans when they are successful and insignificantly lower when they are unsuccessful. However, all of the studies mentioned above investigated the relationship between subordinate race and their superiors' ratings of their performance. This study differed in two respects: the superior was the ratee and not the rater, and the rater and ratee collaborated on a team task with distributed expertise. Perhaps the ratee's leadership position and interaction with the raters influenced the moderated relationship between race and performance ratings.

The results of this study also add to the literature by helping to uncover the cognitive processes underlying the ratings of African American and White American team leaders by their subordinates. Kraiger and Ford (1985) and Dobbins and Platz (1986) suggested that researchers should not simply study whether there is a relationship

between race and performance ratings. They should also investigate why there is a relationship between race and performance ratings. Most of the theories that have subsequently been proposed were not supported by the results of this study. Tokenism, the similar-to-me effect, positivity bias, social identity theory, and intergroup theory all say pretty much the same thing- that African Americans will be rated lower than White Americans because individuals will rate their own race higher. But the results presented here suggest that the cognitive processes involved in rating ones team leader are more involved than simply identifying him as the same race.

In this study, the race of the team leader did not directly affect performance ratings. Instead, the race of the team leader impacted causal attributions, which then influenced performance ratings. As Ilgen and Youtz (1986) hypothesized, the performance of minorities was attributed to different causes than the performance of the majority. But Ilgen and Youtz also thought that good performance would be attributed to external factors for minorities and internal factors for the majority. This was not the case here. Good performance was attributed internally for African American, and not White American, team leaders. However, those attributions did influence performance ratings in the expected direction. Higher internal attributions were related to higher performance ratings and higher external attributions were related to lower performance ratings. The results suggest that African American team leaders are rated differently not simply because the color of their skin is different, but because their performance is attributed to different factors than the performance of White Americans when they are successful.

The results of this study also cast doubt on the applicability of gender research regarding the relationship between team leader gender and performance ratings. A

number of studies have shown that behaviors consistent with a certain gender stereotype are more likely to influence evaluations than behavior that are inconsistent (e.g., Allen & Ebbesen, 1981; Hastie, 1981). Race, in this study, did not operate in a similar fashion. It was the inconsistent behavior of the African American team leader that really influenced his performance ratings. Perhaps consistent and inconsistent behavior elicit different emotional reactions when race, and not gender, is the salient subgroup being stereotyped.

Practical Implications

Aside from the theoretical implications, there are also some practical implications that result from the findings presented in this study. For one thing, it is clear that bias in ratings does not remain constant across ratees. Organizations are not simply using subjective ratings of subordinates by their superiors as performance barometers. The introduction of 360-degree feedback presents the relationship between race and performance ratings from a variety of angles (e.g., Borman, 1974). It is possible that, when minorities are rated by subordinates, the literature regarding the relationship between race and performance ratings is overestimating the bias present in their evaluations. Bias against minorities, as suggested in this study, may only present a problem if those minorities perform poorly.

Another practical implication is that there simply is no practical application at all. Like many of the studies in the literature, the effect sizes presented here are incredibly small (e.g., Pulakos et al., 1989). In fact, the actual performance of the team leader explained much of the variance in performance ratings. Race only had a minor impact on the relationship and effects were only found in the high performance case. It is likely that,

if an organization gathers developmental feedback from a variety of individuals surrounding the ratee, the effects of race on ratings will probably be non-existent.

Study Limitations

This brings me to the limitations of my study, many of which could have influenced the results or lack thereof. As noted above, this study was completed in a laboratory using undergraduate students and not employees. Although every attempt was made to simulate an actual workplace, there were clear differences that could have effected the data. For one thing, the consequences of the students' evaluation of their team leader were not equal to the consequences present in the workplace. The students knew that they were not negatively or positively influencing their leader's life, since he could not get fired, promoted, or demoted. In addition, although they worked together, the team members got paid no matter how well they performed. The bonus of $20 to the top performing subordinates probably did not have the same effect on their emotions as the possibility of getting fired would. Furthermore, the students only worked together for about two hours and only five minutes of that time were spent face-to-face. In the workplace, subordinates would probably work much longer and more interactively with their team leader before ratings his or her performance. Finally, as mentioned earlier, the performance of the team leader was manipulated in this study such that the team performed very well or very poorly. It is likely that the performance of team leaders within the workplace is actually much more moderate than what was created here; certainly the distribution would not be dichotomous. As a result, the strength of the manipulation could have influenced the causal attributions ascribed to the leader's

behavior. For instance, more moderate levels of performance fluctuation could elicit attributions in accordance with the literature on African American stereotypes.

Another limitation of this study has to do with the measurement of causal attributions. Causal attributions are one of the more difficult constructs to measure, which is clearly evidenced by the lack of established measures in the literature. For this study, I attempted to construct a measure tailored to the task of TIDE^2. Although the internal attribution measure evidenced solid internal consistency, convergent, and divergent validity, the external measure was less promising. The alpha for the external scale did not reach .80. And, although the exploratory factor analysis revealed two distinct factors, the confirmatory factor analysis only evidenced a fair fit of the model to the data.

Yet another limitation involves the minority representation in my sample of participants. As the workplace becomes more diversified, minorities will become employed in a variety of jobs across all status levels. Clearly, with a sample of fifteen African Americans, the power to find significant effects utilizing subordinate race was less than optimal.

Future Research

Future research is clearly needed to address some of these limitations. First and foremost is the need to develop a widely applicable measure of causal attributions that can accurately assess whether an individual's behavior is attributed to internal or external causes. The measure of attributions developed by Rusell (1982) was inadequate for its sole use in this study. The internal-external scale, which consisted of three items evidenced poor internal consistency. Part of that was likely due to the number of items.

But another part could also have been due to the ambiguity of the items. The students were often unclear of what exactly the question was asking.

Perhaps an investigation using field data could also improve upon the sample size used in this study. In the low performance condition, African Americans were rated slightly lower than White Americans were, but attributions could not be investigated as potential mediators of a non-significant relationship. Perhaps, with a sample size greater than ninety, the correlation between race and performance ratings would reach significance.

An increased sample size could also introduce subordinate race into the model. In this study, there were only fifteen African Americans. It would be interesting to see whether African Americans viewed their team leaders just as White Americans view their team leaders. Maybe African American subordinates would be less surprised that an African American team leader was performing well than a White American subordinate would.

Other laboratory studies also need to be done to see if, when controlling the level of performance of the subordinates, superiors bias their ratings in a similar fashion to Sackett and DuBois (1991). If so, attribution theory needs to be investigated as a possible influence on that relationship. So far researchers have mostly been proposing that theories such as intergroup theory have the potential to explain why there is bias in ratings of subordinates. But no one has investigated whether attribution theory is the underlying cognitive bias.

References

Alderfer, C.P. (1986). An intergroup perspective on group dynamics. In J. Lorsch (Ed.), *Handbook of organizational behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Allen, R.B., & Ebbesen, E.E. (1981). Cognitive processes in person perception: Retrieval of personality trait and behavioral information. *Journal of Experimental Social Psychology*, 17, 119-141.

Andrews, G.R., & Debus, R.L. (1978). Persistence and causal perceptions of failure: Modifying cognitive attributions. *Journal of Educational Psychology*, 70, 154-166.

Bass, A.R., & Turner, J.N. (1973). Ethnic group differences in relationships among criteria of job performance. *Journal of Applied Psychology*, 57, 101-109.

Bigoness, W.J. (1976). Effect of applicant's sex, race, and performance on employers' performance ratings: Some additional findings. *Journal of Applied Psychology*, 61, 80-84.

Byrne, D. (1971). *The attraction paradigm*. New York: Wiley.

Campion, M.A., Pursell, E.D., & Brown, B.K. (1988). Structured interviewing: Raising the psychometric properties of the employment interview. *Personnel Psychology*, 41, 25-42.

Cardy, R.L., & Dobbins, G.H. (1986). Affect and appraisal accuracy: Liking as an integral dimension in evaluating performance. *Journal of Applied Psychology*, 71, 672-678.

Chapin, M., & Dyck, D.G. (1976). Persistence in children's reading behavior as a function of N length and attribution retraining. *Journal of Abnormal Psychology*, 85, 511-515.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Publishers.

Cox, J.A., & Krumboltz, J.D. (1958). Racial bias in peer ratings of basic airmen. *Sociometry*, 21, 292-299.

Craig, K.M. (1996). Are all newcomers judged similarly? Distinctiveness and time of entry in task-oriented groups. *Small Group Research*, 27, 383-397.

Davidson, M.N., Swigert, S., & Ruderman, M. (1998). A matter of perspective: The effect of race of rater and ratee on managerial performance feedback. *Paper presented at the 1998 Academy of Management Meetings in the Careers Division.*

Davis, L.E., Strube, M.J., & Cheng, L. (1995). Too many African Americans, too many White Americans: Is there a racial balance? *Basic and Applied Social Psychology*, 17, 119-135.

Deaux, K. (1976). Sex: A perspective on the attribution process. In H. Harvey, W.J. Ickes, and R.F. Kidd (Eds.), *New directions in attribution research* (Vol. 1, pp. 1-26). Hillsdale, NJ: Erlbaum.

Deaux, K., & Emswiller, T. (1974). Explanations for successful performance on sex-linked tasks: What is skill for the male is luck for the female. *Journal of Personality and Social Psychology*, 29, 80-85.

DeCarlo, T.E., & Leigh, T.W. (1996). Impact of salesperson attraction on sales managers' attributions and feedback. *Journal of Marketing*, 60, 47-66.

DeJung, J.E., & Kaplan, H. (1962). Some differential effects of race of rater and ratee on early peer ratings of combat aptitude. *Journal of Applied Psychology*, 46, 370-374.

Dobbins, G.H., & Platz, S.J. (1986). Sex differences in leadership: how real are they? *Academy of Management Review*, 11, 118-127.

Doise, W. (1978). *Groups and Individuals: Explanation in social psychology.* Cambridge, England: Cambridge University Press.

Dovidio, J.F., & Gaertner, S.L. (1983). The effects of sex, status, and ability on helping behavior. *Social Psychology Quarterly*, 44, 192-203.

Dweck, C.S. (1975). The role of expectations and attributions in the alleviation of learned helplessness. *Journal of Personality and Social Psychology*, 31, 674-685.

Eaton, N.K., Goer, M.H., Harris, J.H., & Zook, L.M. (1984). *Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1984 fiscal year* (Tech. Rep. No. 660). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Feather, N.T., & Davenport, P. (1981). Unemployment and depressive affect: A motivational and attributional analysis. *Journal of Personality and Social Psychology*, 41, 422-436.

Fox, H., & Lefkowitz, J. (1974). Differential validity: Ethnic group as a moderator in predicting job performance. *Personnel Psychology*, 27, 209-223.

Hamner, W.C., Kim, J.S., Baird, L., & Bigoness, W.J. (1974). Race and sex as determinants of ratings by potential employers in a simulated work-sampling task. *Journal of Applied Psychology*, 59, 705-711.

Hastie, R. (1981). Schematic principles in human memory. In E. Higgens, C. Herman, & M. Zanna (Eds.), *Social cognition: The Ontario symposium.* Hillsdale, N.J.: Erlbaum.

Heider, F. (1958). *The psychology of interpersonal relations.* New York: Wiley.

Heilman, M.E. (1983). Sex bias in work settings: The lack of fit model. In L.L. Cummings & B.M. Straw (Eds.), *Research in organizational behavior* (Vol. 5, pp. 269-298). Greenwich, CT: JAI Press.

Hollenbeck, J.R., Ilgen, D.R., Sego, D.J., Hedlund, J., Major, D.A., & Phillips, J. (1995). Team decision making accuracy under difficult conditions: Construct validation of potential manipulations using the TIDE2 simulation. In M.T. Brannick, E. Salas, and C. Prince (Eds.), *Team Performance, assessment and measurement: Theory, research and applications*. Hillsdale, NJ: Erlbaum.

Hunter, J.E., & Hirsch, H.R. (1987). Applications of meta-analysis. In C.L. Cooper and I.T. Robertson (Eds.), *International review for industrial and organizational psychology*. New York: Wiley.

Ickes, W., & Layden, M.A. (1978). Attributional styles. In J. Harvey, W. Ickes, & R. Kidd (Eds.), *New directions in attributional research* (Vol. 2, pp.119-152). Hillsdale, NJ: Lawrence Erlbaum.

Ilgen, D.R. (1994, April). *Jobs and roles: Accepting and coping with changing structure in organizations*. Paper presented at the Army Research Institute Selection and Placement Conference, Alexandria, VA.

Ilgen, D.R., & Youtz, M.A. (1986). Factors affecting the evaluation and development of minorities in organizations. In K. Rowland & G. Ferris (Eds.), *Research in personnel and human resource management: A research manual* (pp. 307-337). Greenwich, CT: JAI Press.

Jackson, L.A., Sullivan, L.A., & Hodge, C.N. (1993). Stereotype effects on attributions, predictions and evaluations: No two social judgements are quite alike. *Journal of Personality and Social Psychology*, 65, 69-84.

Jensen, A.R. (1980). Bias in mental testing. New York: Free Press.

Jones, E.E., & McGillis, D. (1976). Correspondent inferences and the attribution cube: A comparative reappraisal. In J.H. Harvey, W.J. Ickes, & R.F. Kidd (Eds.), *New directions in attribution research* (Vol. 1, pp. 389-420). Hillsdale, NJ: Erlbaum.

Jones, E.E., & Sigall, H. (1971). The Bogus Pipeline: A new paradigm for measuring affect and attitudes. *Psychological Bulletin*, 76, 349-364.

Kanter, R.M. (1977). *Men and women of the corporation*. New York: Basic Books.

Kelley, H.H. (1971). *Attributions in social interactions*. Morristown, NJ: General Learning Press.

Kraiger, K., & Ford, K.J. (1985). A meta-analysis of ratee race effects in performance ratings. *Journal of Applied Psychology*, 70, 56-65.

Kraiger, K., & Ford, J.K. (1990). The relation of job knowledge, job performance, and supervisory ratings as a function of ratee race. *Human Performance*, 3(4), 269-279.

Landy, F.J., & Farr, J.L. (1980). Performance Rating. *Psychological Bulletin*, 87, 72-107.

Lin, T.R., Dobbins, G.H., & Farth, J.L. (1992). A field study of race and age similarity effects on interview ratings in conventional and situational interviews. *Journal of Applied Psychology*, 77, 363-371.

Martinko, M.J., & Gardner, W.L. (1987). The leader/member attribution process. *Academy of Management Review*, 12, 235-249.

Messick, D., & Mackie, D. (1989). Intergroup relations. *Annual Review of Psychology*, 40, 45-81.

Mitchell, T.R., & Wood, R.E. (1980). Supervisor's responses to subordinate's poor performance: A test of an attributional model. *Organizational Behavior and Human Performance*, 25, 123-128.

Mount, M.K., Sytsma, M.R., Hazucha, J.F., & Holt, K.E. (1997). Rater-ratee race effects in developmental performance ratings of managers. *Personnel Psychology*, 50, 51-69.

Nieva, V.F., & Gutek, B.A. (1980). Sex effects on evaluation. *Academy of Management Journal*, 5, 267-276.

Norvell, N., & Forsyth, D.R. (1984). The impact of inhibiting or facilitating causal factors on group members' reactions after success and failure. *Social Psychology Quarterly*, 47, 293-297.

O'Leary, V.E., & Hansen, R.D. (1984). Sex as an attributional fact. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (pp. 133-177). Lincoln: University of Nebraska.

Pettigrew, T.F. (1979). The ultimate attribution error: Extending Allport's cognitive analysis of prejudice. *Personality and Social Psychology Bulletin*, 5, 461-476.

Phillips, J.M. (1997). Antecedents and consequences of leader utilization of staff information in decision making teams: Addressing a leadership dilemma. *Doctoral dissertation*, Michigan State University, Department of Psychology, East Lansing, MI.

Prewett-Livingston, A.J., Field, H.S., Veres, J.G., III, & Lewis, P.M. (1996). Effects of race on interview ratings in a situational panel interview. *Journal of Applied Psychology*, 81, 178-186.

Pulakos, E.D., White, L.A., Oppler, S.H., & Borman, W.C. (1989). Examination of Race and Sex Effects On Performance Ratings. *Journal of Applied Psychology*,74, 770-780.

Rand, T.M., & Wexley, K.N. (1975). Demonstration of the effect "similar to me" in simulated employment interviews. *Psychological Reports*, 36, 535-544.

Rice, R.W., Bender, L.R., & Vitters, A.G. (1980). Leader sex, follower attitudes toward women, and leadership effectiveness: A laboratory experiment. *Organizational Behavior and Human Performance*, 25, 46-78.

Rosen, B., & Jerdee, T.(1973). The influence of sex-role stereotypes on evaluations of male and female supervisory behavior. *Journal of Applied Psychology*, 57, 44-48.

Rosenbaum, R.M. (1972). *A dimensional analysis of the perceived causes of success and failure*. Unpublished doctoral dissertation, University of California, Los Angeles.

Rotter, J.B. (1966). Generalized expectancies for internal-stable versus external-unstable control of reinforcement. Psychological Monograph, 80, 1-28.

Rotter, N., & Rotter, G.S. (1969, April). Race, work performance and merit rating: An experimental evaluation. *Paper presented at the meeting of the Eastern Psychological Association, Philadelphia.*

Russell, D.W. (1982). The causal dimension scale: A measure of how individuals perceive causes. *Journal of Personality and Social Psychology*, 42, 1137-1145.

Russell, D.W., McAuley, E., & Tarico, V. (1987). Measuring causal attributions for success and failure: A comparison of methodologies for assessing causal dimensions. *Journal of Personality and Social Psychology, 42*, 1248-1257.

Sackett, P.R., & DuBois, C.L.Z. (1991). Rater-ratee race effects on performance evaluation: Challenging meta-analytic conclusions. *Journal of Applied Psychology, 76, 873-877.*

Sackett, P.R., DuBois, C.L.Z., & Noe, A.W. (1991). Tokenism in performance evaluation: The effects of work group representation on male-female and White American-African American Differences in performance ratings. *Journal of Applied Psychology*, 76, 263-267.

Schachter, S., & Singer, J.E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 69, 379-399.

Schmidt, F.L., & Johnson, R.H. (1973). Effect of race on peer ratings in an industrial setting. *Journal of Applied Psychology*, 57, 237-241.

Schmitt, N., & Hill, T.E. (1977). Sex and race composition of assessment center groups as a determinant of peer and assessor ratings. *Journal of Applied Psychology*, 62, 261-264.

Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology*, 82, 719-730.

Silver, W.S., Mitchell, T.R., & Gist, M.E. (1995). Responses to successful and unsuccessful performance: The moderating effect of self-efficacy on the relationship between performance and attributions. *Organizational Behavior and Human Decision Processes, 62*, 286-299.

Smith, T.W. (1990). *Ethnic images* (GSS Topical Report No. 19). Chicago: University of Chicago, National Opinion Research Center.

Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology.* Cambridge, England: Cambridge University Press.

Taylor, S.E. (1981). A categorization approach to stereotyping. In D.L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 83-114). Hillsdale, NJ: Erlbaum.

Turner, J.C. (1984). Social identification and psychological group formation. In H.Tajfel (Ed.), *The social dimension: European developments in social psychology* (pp. 518-538). Cambridge, England: Cambridge University Press.

Turner, J.C., Hogg, M.A., Oakes, P.J., Reicher, S.D., & Wetherell, M.S. (1987). *Rediscovering the social group: A self-categorization theory.* New York: Basil African Americanwell.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92, 548-573

Wendelken, D.J., & Inn, A. (1981). Nonperformance influences on performance evaluations: A laboratory phenomenon? *Journal of Applied Psychology*, 66, 149-158.

Whitehead, G.I. III, Smith, S.H., & Eichhorn, J.A. (1982). The effect of subject's race and other's race on judgments of causality for success and failure. *Journal of Personality*, 50, 193-202.

Wilder, D.A. (1986). Social categorization: Implications for creation and reduction of intergroup bias. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 291-355). San Diego, CA: Academic Press.

Wyer, R.S., & Scrull, T.K. (1980). The processing of social stimulus information: A conceptual integration. In R. Hastie et al. (Eds.), *Person memory: The cognitive basis of social perception*. Hillsdale, N.J.: Erlbaum.

Ziller, R.C., Behringer, R.D., & Goodchilds, J.D. (1960). The minority newcomer in open and closed groups. *Journal of Psychology*, 50, 75-84.

Zuckerman, M. (1979). Attribution of success and failure revisited, or the motivational bias is alive and well in attribution theory. *Journal of Personality*, 47, 245-287.

APPENDIX A

Items Measuring Internal and External Attributions

5- Strongly agree
4- Agree
3- Neither Agree nor Disagree
2- Disagree
1- Strongly Disagree

i1. Our team leader is (not) very smart, and that contributed to our performance.
i2. Alpha's (low) high intelligence greatly affected our score.
i3. Our team leader (did not) put a lot of effort into the game, and that influenced our performance.
i4. Our team leader worked (not) very hard to accurately assess all nine attributes.
e1. Our team leader was (not) very lucky on many of the trials, and that affected our average score.
e2. Our performance was due to a streak of (bad) luck by Alpha.
e3. This game was (not) easy, and that contributed significantly to our team leader's performance.
e4. It was (not) easy to determine the threat of the targets, and that is why our team leader did (not) do so well.

APPENDIX B

Performance Rating Items

5- Strongly agree
4- Agree
3- Neither Agree nor Disagree
2- Disagree
1- Strongly Disagree

1. Our team leader performed very well on this task.
2. Without our team leader, we would have had little chance at earning the bonus money.
3. Our team leader was skilled at sorting all nine attributes into the three rules (i.e., category, location, and motion).
4. Our team leader could have earned the bonus money without my input.
5. The number of hits we got was mostly due to Alpha.
6. Alpha was a good team leader.
7. Our team leader was very good at combining the attributes within each of the rules.
8. Alpha could not have performed any better.
9. If I were the team leader, I could not have done any better.
10. Alpha was very good at quickly measuring all nine attributes.
11. Alpha was very good at comparing the general values on his/her individual sheet with the values in his/her measure summary box.
12. Our team leader made this game look easy.
13. Our team leader was very good at combining the rules into one final team decision.
14. Our team leader did **not** measure all nine attributes quick enough.
15. Alpha did **not** perform very well on this task.
16. Alpha had difficulty combining the three rules into one overall team decision.
17. Alpha made a lot of mistakes during the game.
18. Alpha did **not** remember that one "safe" attribute meant that the entire rule was safe.
19. Alpha could have done a lot better.

APPENDIX C

Performance Rating Items Scaled 1 to 100

1. On a scale of 1 to 100, with 100 being perfect performance, how well do you think Alpha performed?

2. On a scale of 1 to 100, how good of a team leader was Alpha?

APPENDIX D

Planning for Participants: The Power Analysis

Planning for power is an effective tool that can, if utilized properly, reduce the probability of making a Type II error when conducting research. Statistical power is defined as the probability of correctly rejecting the null hypothesis given (a) the sample size, (b) the effect size, (c) alpha, and (d) the power of a statistical test. Power, for this study, was set at the conventional .80 for each of the hypotheses.

Hypothesis 1. In a meta-analysis completed by Schmitt et al. (1997), the population correlation between race and performance ratings was estimated to be .18. Using Cohen's (1988) power tables, 160 subjects would be needed to achieve a power of .80.

Hypothesis 2. Hamner, Kim, Baird, and Bigoness (1974) found that thirty percent of the variance in performance ratings was due to the individual's actual level of performance. Using this estimate of r (.50 to be conservative), and the power tables from Cohen, 23 subjects would be needed to achieve a .80 power level.

Hypotheses 3 & 4. Whitehead et al. (1982) examined the three-way interaction between ratee race, rater race, and actual performance on performance attributions. Their results evidenced an interaction effect $F(4,170)=2.39$ $p<.052$, which was on the small side. For a small effect size, I would need 700 subjects for a power of .80 according to the tables in Cohen. For 200 subjects, my power level would be .41. But, since my hypothesis is only a two-way interaction, my effect size should be larger, thereby reducing the amount of subjects that I would need. For a small-to-moderate effect size ($R^2=.04$), 160 subjects would be enough for a power level of .80. Unfortunately my literature search did not turn up any articles with relevant two-way interactions.

Hypotheses 5 & 6. Since my final hypothesis has not, as of yet, been investigated, I must infer correlation values based on either a small or a moderate effect size. For a small

effect size (R^2=.01), I would need 700 subjects for a power of .80. For 200 subjects, my power level would be .41. For a small-to-moderate effect size (R^2=.04), 160 subjects would be enough for a power level of .80.

Therefore, given the assumptions in this analysis, a minimum sample size of 160 is thought to be adequate to test the hypotheses.