

**THE TWO-WAY GENDER BIAS IN  
MANAGEMENT EVALUATIONS AND DECISION MAKING:**

Evaluations of Managers vs. Evaluations of Grievants

by

Harsh K. Luthar


Dissertation submitted to the Faculty of the  
Virginia Polytechnic Institute and State University  
in partial fulfillment of requirements for the degree of

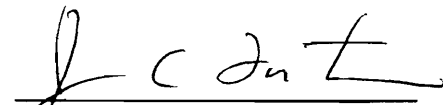
DOCTOR OF PHILOSOPHY

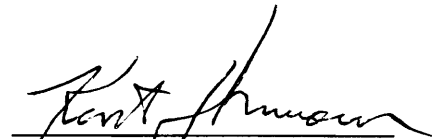
in

Management

APPROVED:

  
Dr. T.W. Bonham, Chairman

  
Dr. Jim C. Fortune

  
Dr. Kent F. Murrmann

  
Dr. Dane M. Partridge

  
Dr. K. Dow Scott

August 20, 1993

Blacksburg, Virginia

**THE TWO-WAY GENDER BIAS IN  
MANAGEMENT EVALUATIONS AND DECISION MAKING:**

Evaluations of Managers  
v.  
Evaluations of Grievants

by

Harsh K. Luthar

Dr. T.W. Bonham, Chairman  
Management

(ABSTRACT)

This study consisted of two field experiments and examined how managerial evaluations may be influenced by the gender and gender-role behavior of the person being evaluated. Literature streams in the domains of leadership, performance evaluation, grievance arbitration, and criminology were reviewed and integrated to derive a conceptual framework. The research hypotheses focused on how male and female managers would differ in their evaluations of male and female target persons in two different contexts.

Two cases, each approximately three pages long, were written for the purpose of this study. In both cases, the gender and gender-role behavior of the target person were manipulated. A pilot study, in which 255 students responded to the case of the grievant and 290 students responded to the case of the manager, was carried out to test the psychometric properties of the scales as well as the integrity of the gender-role manipulation. The results indicated the success of the manipulations. The evaluation measures of interest were

found to have high reliabilities. In the actual study, research subjects consisted of 129 unit directors or higher level administrators working for the Virginia Cooperative Extension. The subjects were asked to respond to both cases and evaluate the target persons. The sequence in which the cases were presented was built into the design of the study.

There were no significant findings pertaining to the proposed hypotheses. Other results indicated that in the grievance resolution context, female unit directors rated aggressive male workers lower than aggressive female workers on performance and abilities. Relative to male unit directors, female unit directors gave aggressive lower-level male workers unfavorable evaluations. In the context of evaluating managers, female unit directors relative to male unit directors more favorably evaluated the democratic manager. The difference between male and female unit directors was larger when they evaluated the same-sex democratic manager. Female unit directors gave the democratic female manager higher evaluations than those given to the democratic male manager by male unit directors.

The discussion focuses on the implications of results for conceptual and theoretical development in the gender bias literature, as well as what the results may mean for organizations and practicing managers. The potential limitations of the study are identified, and the study is concluded with suggestions for future research.

## ACKNOWLEDGEMENTS

I feel fortunate to have had Dr. T.W. Bonham as my chair who guided this project from the conceptual development to the data analyses and interpretation stage. Dr. Bonham's advice and assistance throughout the different phases of this dissertation were invaluable. He could be very "hands on" when needed (such as in helping me understand factor analysis from the computer printouts) but at the same time gave me the freedom (and forced me) to confront the tough research questions and think for myself.

Dr. Bonham was instrumental in keeping me focused on the goal, when perhaps if left to my own devices, I may have gone off-track. In doing this dissertation with him, I started to mature as a researcher, became aware of the intricacies of the research process, and developed confidence in my ability to make judgments and decisions which are necessary at various junctures of a project.

I express to Dr. Bonham my sincere appreciation and gratitude for all that he has done for me. He is a true master of the research process and takes very seriously the job of developing his Ph.D students. He is highly accessible to graduate students (not just the ones he chairs) and often goes out of his way to help them. I will always remember his various unstated mottos (on occasion made explicit in

different ways) which were reflected in suggestions such as "Think about what you are doing" "Get to know your data" "Don't write 10 more pages when I ask you to revise something" "Get to work, don't waste time, and get the job done." Thank you Dr. Bonham for all your help. I look forward to continue working with you and hope to benefit from your counsel and insight in the future.

Next, I thank my dissertation committee members (in alphabetical order) Dr. Jim C. Fortune, Dr. Kent F. Murrmann, Dr. Dane M. Partridge and Dr. K. Dow Scott. Their constant encouragement, kindness, and good humor made the Ph.D program bearable.

I would like to note that over the years, Dr. Fortune has been an indispensable resource person for me in the area of measurement and methods, and we have spent much time in his office talking about scaling issues faced by researchers. Because of Dr. Fortune's encouragement, I have developed a serious interest in the area of measurement and research methods, and hope at some point to make a contribution in this field. It has been a real joy having him on the committee and working with him.

I thank Dr. Kent Murrmann for his insights, moral support and good humor over the years. Dr. Murrmann has often said to me, and probably to other Ph.D students, "These are the best years of your life." While it is rather difficult to agree

with him on that, I must admit it always brings a smile to my face when I think of him saying that with such enthusiasm and sincerity. Although it is in his nature to keep a low profile, Dr. Murrmann is a genuine scholar of remarkable depth and a person of great insight. I feel very privileged having had Dr. Murrmann on my committee.

I thank Dr. Dane Partridge for his friendship and help over the last four years. I feel particularly fond of Dr. Partridge as he (along with Dr. Murrmann) has told me over the years what a fine writer I am. I have always secretly enjoyed hearing that (while pretending to be modest), particularly coming from Dr. Partridge who has a reputation for being an outstanding writer and whom I hold in high esteem for his scholarship. I certainly look forward to a rewarding personal and a professional relationship with Dr. Partridge in the future.

I thank Dr. Dow Scott for his excellent advice and suggestions made to me during the dissertation process. Dr. Scott made a valuable suggestion to me when he remarked that I think about the unit directors as a potential sample for my dissertation. His generally cheerful disposition and good humor make him highly approachable, and I have always enjoyed and learned from our conversations. Over the years, I have become aware that Dr. Scott is not only a serious and prolific researcher, but also an honorable individual with a good

heart. I feel very proud to have had Dr. Scott on my committee.

I would like to also thank Dr. Robert B. Frary and his office staff for their valuable assistance in scanning the opscan forms for the pilot study and the study with the unit directors. Dr. Frary was particularly generous and always made time to answer questions pertaining to SAS procedures during the pilot stage of the project.

I would like to thank Dr. Robert J. Litschert, the director of graduate students, and Dr. Jon M. Shepard, chairman of the management department. They have provided the Ph.D students with resources and much needed support over the years. Their decision to give Ph.D students their own offices as well as a computer room with Desk Jet printers was helpful to many of us who needed a place to work on campus. Dr. Litschert's concern for graduate students and his generosity towards them is well known and I thank him for his help and support over the last four years. Let me also note that Dr. Shepard, for whom I worked as a G.A during my last year, gave me only little work which allowed me more time to devote to the dissertation. All things add up and help.

I would like to acknowledge the faculty members from whom I have had the good fortune to take various classes and seminars. They have my heart felt thanks for creating a learning environment conducive to scholarship as well as

collegiality. They are (in alphabetical order): Dr. T.W. Bonham (Organization Behavior), Dr. Anthony T. Cobb (Organization Behavior), Dr. Jim C. Fortune (Methods and Statistics courses), Dr. James C. Impara (Statistics), Dr. Richard G. Krutchkoff (Statistics), Dr. Robert M. Madigan (Human Resources), Dr. Kent B. Monroe (Methods), Dr. Kent F. Murrmann (Human Resources/Labor), Dr. Dane M. Partridge (Labor), Dr. Jerald F. Robinson (Labor), and Dr. K. Dow Scott (Human Resources).

I would like to thank my many friends and colleagues in the Ph.D program with whom I have had classes and seminars for their understanding, moral support, comradrie and collegiality over the last 4 years. In particular, I will remember Doug Bolon, Jim Bishop, Tony Townsend, Beverly Little, and Bob D'Intino fondly. We have shared our joys and sorrows in the Ph.D program and have shared many a laughter. (Laughter which was occasionally uncontrollable, went on for long periods until it made our stomachs hurt and our eyes tear.) I wish my friends well in their academic careers.

I would like to thank my family for their support through this long struggle. First I acknowledge my parents, Raj and Rajendra Luthar, for the pivotal role they have played in my academic career. They made me aware of the importance of developing a love for learning and seeking knowledge long time ago. Because of the seeds they sowed, a desire to seek the



truth remains strong within me and guides my endeavors. Second, I would like to thank my wife Anshu and myself for holding everything together during the years that we have both been students and creating a reasonably normal environment for our son to grow up in. Finally, a special thanks to my son Jay who, despite having sacrificed much play time with his father over the last four years, remains steadfast in his affections. His enthusiasm and energy have been and will be a vitalizing influence for me.

**TABLE OF CONTENTS** **PAGE**

**CHAPTER 1 : INTRODUCTION**

BACKGROUND.....	1
PURPOSE OF THE STUDY.....	4
SIGNIFICANCE OF THE STUDY.....	6

**CHAPTER 2 : LITERATURE REVIEW**

INTRODUCTION.....	10
<b>GENDER BIAS IN EVALUATION</b>	
OF PERFORMANCE AND LEADERSHIP.....	11
The Price Waterhouse Case.....	11
The Goldberg Study	
and its Replications.....	16
The Role of Organizational	
Social Composition.....	20
Overvaluation of	
Performance of Women.....	24
Gender and Evaluations	
of Managers.....	27

<b>GENDER BIAS IN DISPENSATION OF JUSTICE.....</b>	<b>31</b>
Gender Bias Models from Criminology.....	31
Gender Biases in Management	
Evaluations of grievances.....	36
Gender Biases in Arbitration.....	40
Implications for Management	
Evaluations and Decisions.....	43

#### **THEORY AND HYPOTHESES**

<b>AN INTEGRATIVE AND THEORETICAL ANALYSIS.....</b>	<b>45</b>
The Nature of Sex Stereotypes.....	46
Gender Based Expectations.....	48
The Role of Organizational Variables.....	51
Causal Attributions and Gender Bias.....	55
The Two-Way Gender Bias.....	60
Male v. Female Decision Makers.....	64
Research Hypotheses.....	66

### **CHAPTER 3 : METHODOLOGY**

<b>INTRODUCTION.....</b>	<b>68</b>
--------------------------	-----------

#### **COMPARING FIELD STUDIES**

<b>AND FIELD EXPERIMENTS.....</b>	<b>68</b>
-----------------------------------	-----------

    Advantages and Disadvantages of

Field Studies.....	69
--------------------	----

Advantages and Disadvantages of Experimental Methodology.....	70
Appropriate Design for the Study.....	73
<b>CHOICE OF DATA COLLECTION METHOD AND SAMPLE.....</b>	<b>73</b>
Field Experiment	
Using the Mail Questionnaire.....	75
On-Site Study in an Organization.....	76
On-Site Study at a Management Training Center.....	78
Method of Data Collection Chosen.....	79
<b>DEVELOPMENT OF MEASURES AND PRE-TESTING.....</b>	<b>79</b>
Summary Description.....	84
Theoretical Justification for the Scale.....	85
Choice and Construction of the Dependent Variables.....	92
<b>PILOT STUDY.....</b>	<b>95</b>
Reliability Issues.....	96
The Sample.....	98
Manipulations.....	100
Design and Procedures.....	101

The Nature and Quality of Pilot Study Data.....	105
General Analysis for the Pilot Study.....	108
<b>RESULTS FROM THE GRIEVANT SCENARIO.....</b>	<b>113</b>
Constructing the Dependent Variables.....	116
Descriptive Statistics.....	119
Statistical Analyses for Manipulation Checks.....	124
Conclusions.....	127
<b>RESULTS FROM THE MANAGER SCENARIO.....</b>	<b>130</b>
Constructing the dependent variables.....	131
Descriptive Statistics.....	135
Statistical Analyses for Manipulation Checks.....	138
Conclusions.....	139
<b>REFINING THE DESIGN BASED ON THE RESULTS OF THE PILOT.....</b>	<b>141</b>

<b>CHOICE OF THE SITE, PLANNED ANALYSIS AND DATA COLLECTION.....</b>	<b>143</b>
The Theoretical and Practical Implications of the Sample.....	145
Planned Statistical Analysis to Test the hypotheses.....	146
The Influence of Cell Size on the Analysis and Interpretation.....	148
Data Collection Procedures.....	152
 <b>CHAPTER 4 : RESULTS</b>	
<b>INTRODUCTION.....</b>	<b>157</b>
 <b>SAMPLE CHARACTERISTICS AND MEASUREMENT.....</b>	<b>157</b>
Subjects' Response Rate.....	159
Missing Values in the Data.....	160
Constructing Measures Using Factor Analysis.....	161
 <b>RESULTS FROM THE GRIEVANT SCENARIO.....</b>	<b>163</b>
Choosing the Dependent Variable for Analysis.....	167
Descriptive Statistics.....	169
Testing the Hypotheses.....	173

<b>RESULTS FROM THE MANAGER SCENARIO.....</b>	<b>177</b>
Choosing the Dependent Variable	
for Analysis.....	180
Descriptive Statistics.....	182
Testing the Hypotheses.....	186
<b>SUMMARY.....</b>	<b>189</b>

## **CHAPTER 5 : DISCUSSION**

<b>INTRODUCTION.....</b>	<b>190</b>
Uniqueness of the Study.....	190
<b>DISCUSSION-THE GRIEVANT SCENARIO.....</b>	<b>191</b>
The Effectiveness of the	
Gender Role Manipulation.....	192
How the Results fit into the Literature.....	198
Questioning the	
"Chivalry/Paternalism" Effect.....	202
Would the Chivalry/Paternalism Effect Differ	
Between Managers and Arbitrators.....	207
Summary of the Discussion-Grievant Case.....	213
<b>DISCUSSION-THE MANAGER SCENARIO.....</b>	<b>214</b>
The Strength of the	
Gender Role Manipulation.....	215

The Amount of Information	
Given in the Case.....	219
The Role of the	
Organizational Composition.....	221
Summary of the Discussion--	
Case of the Manager.....	222
<b>CONCLUSIONS AND IMPLICATIONS.....</b>	<b>224</b>
<b>MAKING SENSE OF IT ALL.....</b>	<b>225</b>
Having a Female Boss	
versus a Male Boss.....	226
Possible Consequences.....	227
Preferred Managerial Style.....	230
<b>LIMITATIONS OF THE STUDY.....</b>	<b>232</b>
Internal Validity Considerations.....	233
External Validity Considerations.....	236
Construct Validity Considerations.....	239
Statistical Conclusion Validity	
Considerations.....	242
<b>SUGGESTION FOR FUTURE RESEARCH.....</b>	<b>244</b>
Conclusion.....	247



**REFERENCES.....248**

**APPENDIX.....263**

CASES.....264

SCALING INSTRUCTIONS.....292

DEMOGRAPHIC QUESTIONNAIRE.....293

RESUME.....294

**TABLES****PAGE****TABLE 1--Levels of Independent Variables**

Grievant Scenario-Pilot.....119

**TABLE 2--Means, Standard Deviations, and Alphas  
of Factor-Based Scales**

Grievant Scenario-Pilot.....120

**TABLE 3--Correlations**

Grievant Scenario-Pilot.....121

**TABLE 4--Levels of Independent Variables**

Manager Scenario-Pilot.....135

**TABLE 5--Means, Standard Deviations, and Alphas  
of Factor-Based Scales**

Manager Scenario-Pilot.....136

**TABLE 6--Correlations**

Manager Scenario-Pilot.....137

**TABLE 7--Levels of Independent Variables**

Grievant Scenario-Unit Directors.....169

<b>TABLE 8--Means, Standard Deviations, and Alphas</b> of Factor-Based Scales Grievant Scenario-Unit Directors.....	170
<b>TABLE 9--Correlations</b> Grievant Scenario-Unit Directors.....	171
<b>TABLE 10--Mean Ratings of the Targets Person by Unit Directors</b> on the Dependent Variable DECFAIR/PERFORM Grievant Scenario-Unit Directors.....	174
<b>TABLE 11--Levels of Independent Variables</b> Manager Scenario-Unit Directors.....	182
<b>TABLE 12--Mean, Standard Deviations, and Alphas</b> of Factor-Based Scales Manager Scenario-Unit Directors.....	183
<b>TABLE 13--Correlations</b> Manager Scenario-Unit Directors.....	184
<b>TABLE 14--Mean Ratings of the Target Persons by</b> Unit Directors on the Dependent Variable GENERAL Manager Scenario-Unit Directors.....	187

## CHAPTER 1

### INTRODUCTION

#### BACKGROUND

A discussion of gender issues has been prevalent in the popular media for the last two decades and shows no sign of abating in the near future. Much research today is being conducted and published, in both practitioner and academic journals in the field of management, that deals with topics such as gender and power in organizations (Ragins & Sundstrom, 1989), sexual harassment in the workplace (Terpstra & Baker, 1988, 1992), gender differences in trust of supervisors (Scott, 1983), gender bias in evaluation of managers (Rosen & Jerdee, 1973; Bartol & Butterfield, 1976; Brown & Gies, 1984; Eagly & Johnson, 1990; Eagly, Makhijani, & Klonsky, 1992), gender bias in performance evaluation (Nieva & Gutek, 1980; Swim, Borgida, Maruyama, & Myers, 1989; Top, 1991), gender and workplace justice (Dalton & Todor, 1985a, 1985b; Dalton, Todor, & Owen, 1987; Dalton, Mesch, Owen, & Todor, 1992), gender biases in arbitration (Bemmels, 1988a, 1988b, 1988c, 1990a, 1990b, 1991a, 1991b; Caudil & Oswald, 1991, 1992), gender differences in starting salaries (Gerhart, 1990), career progression of male and female managers (Cannings & Montmarquette, 1991; Stroh, Brett & Reily, 1992), and differential responses of male and female managers toward poor performers in the organization (Dobbins, 1985).

The interest in the role of women at work is understandable for many reasons. First, women have been increasingly joining the work force in recent decades. According to projections of the Bureau of Labor Statistics (BLS), by the year 2000, women will constitute almost half of the total work force. Even today, many serve in key positions as professionals and managers in both private and public organizations.

Second, it is clear that women face certain special challenges in the workplace. Women are often not accepted as equals by their male peers, and in their roles as managers, they may at times face resistance from male subordinates (Harlan & Weiss, 1982; Bayes & Newton, 1978; Mayes, 1979). Women may be perceived as not having the emotional maturity, experience, drive, and other qualities required to be successful as managers (Bowman, Wortney, & Greyser, 1965; Basil and Traver, 1972; Schein, 1973). Even very recent studies indicate that male managers perceive men in general to be more similar to successful managers than women in general (see for example, Heilman, Block, Simon, & Martell, 1989).

Third, although the law provides for giving women equal opportunity in employment, there is much evidence that sex discrimination, sexual harassment and other forms of gender bias continue to play a role in the careers of women (Terpstra & Baker, 1988, 1992). The promotion of women to higher management positions may be perceived as being due more to

affirmative action efforts of the organization than to merit. This may lead to a devaluation of the performance of the female manager (Jacobson & Koch, 1977; Summers, 1991).

The fact that there is an overwhelming interest in our society in gender issues at the workplace became evident during the hearings (1991) on the Clarence Thomas nomination to the Supreme Court. The nation sat transfixed by the television as Professor Anita Hill charged that Judge Thomas, her former boss, had sexually harassed her on the job. Even more recently, the United States Navy was rocked by a sexual harassment scandal that led to the resignations and reassignment of several top Navy officials. Acting Navy Secretary Sean O' Keefe conceded that the Navy suffered from a serious "cultural problem which has allowed demeaning behavior and attitudes towards women to exist" (*Wall Street Journal*, September 25, 1992: A10). Such widely publicized incidents have served to heighten national sensitivity with regard to the difficulties faced today by working women in diverse domains.

Clearly, as women increase in numbers as well as status at the workplace, it should be expected that issues pertaining to women will become increasingly critical for organizations to address and for scholars to research. This is particularly important in view of the fact that even relatively recent research results indicate that gender stereotypes remain deeply rooted in our society (e.g. Deaux & Lewis, 1983;

Martin, 1986; Heilman, Block, Simon, & Martell, 1989; Powell & Butterfield, 1989). Recognizing the reality of this phenomenon, a plurality of Justices in the 1989 Supreme Court decision of Price Waterhouse v. Hopkins endorsed the concept of sex stereotyping to be legally relevant to sex discrimination litigation.

### PURPOSE OF THE STUDY

For reasons cited above, this research seeks to make a contribution to the literature on gender bias. The proposed study is unusual in that a serious attempt is made to integrate results from many different areas of research such as leadership evaluation, performance evaluation, grievance arbitration, and criminology. This allows for a richer set of hypotheses to be tested for gender bias can be conceived of as a phenomenon that may work in opposing directions--sometimes against women and sometimes to the advantage of women.

Traditionally, gender bias has been thought of as something that can only have a detrimental impact on women in their roles as workers and managers. However, a review of the relevant literature demonstrates that negative evaluations of women are not always the rule in organizations. In fact, in some instances, decision makers may favorably evaluate women over similarly situated men. Such findings are difficult to reconcile with conventional assumptions about gender bias. For purposes of clarity in this context, the conceptualization

that bias can work both ways for women and lead to decisions that may be favorable or unfavorable to them will be referred to as the "two-way gender bias."

The idea that gender bias at times may work in favor of women is not original (see for example, Deaux & Taynor, 1973; Taynor & Deaux, 1973; Hamner, Kim, Baird, & Bigoness; 1976; Bigoness, 1976; Nieva and Gutek, 1980;) and has some empirical support. However, researchers have rarely explored the possibility of a two-way gender bias existing in individuals and manifesting itself in different directions depending on the context and the nature of the decision to be made.

A scanning of the organizational literature on gender bias indicates little serious discussion on how to theoretically interpret divergent results found in a variety of evaluation settings. Heilman, Martell, and Simon (1988), recognizing the importance of this issue, stated, "It appears, for example, that sometimes women may actually benefit from their sex rather than be disadvantaged by it. Few attempts have been made to explain this finding or to reconcile it with current theories of sex bias" (P.99). Indeed, studies examining the behavior of decision makers in different evaluation contexts in organizations, allowing for the possibility of a two-way gender bias, are virtually non-existent.

Because of the lack of empirical literature in the area, particularly within the framework of frequently occurring



evaluation decisions that have to be made in organizations, the role and importance of the two-way gender bias remains unclear. This study seeks to fill this gap by investigating how decision makers are influenced by gender in their evaluations of managers and workers in two different contexts. Specifically, the central research question is this: Do managerial judgements and decisions vary as a function of the gender of the manager making the decision, the gender of the person being evaluated, the extent to which the behavior of the individual being evaluated has been gender-role congruent, and whether the situation involves evaluating a manager or a grievant. The purpose of the study is to investigate differences in evaluations and outcomes associated with the possibility of a two-way gender bias among decision makers and to explore their implications for organizations.

#### SIGNIFICANCE OF THE STUDY

This study is significant because it is the first attempt to link two relatively diverse streams of literature and bring them under one theoretical umbrella. Although numerous studies have investigated gender bias in the context of leadership evaluation, performance evaluation, and grievance arbitration, no conceptual framework has been proposed in the organizational literature that would explain the seemingly disparate and often inconsistent findings in these areas. The lack of an overarching theoretical framework makes it

difficult to use results from one area of research on gender bias to understand similar phenomena in a different context.

It is even problematic to generalize results within the same field. For example, it is currently unclear whether the results from the arbitration literature on gender bias have any implications for grievance resolution in organizations where unions are not present to assist in grievance handling. Drawing inferences about non-union organizations is further complicated by the fact that studies directly investigating gender bias in the grievance resolution context, using managers as subjects, are extremely rare.

More than a decade ago Nivea and Gutek (1980) offered a comprehensive review of the literature pertaining to sex effects in evaluation. In recent times, as the number of published studies has grown, researchers have often relied on meta-analysis techniques to summarize findings in the gender bias literature. Such reviews necessarily must focus very narrowly on studies which are very similar to one another. This has led to the field becoming more and more specialized, as researchers choose to focus on a particular area of expertise, such as leadership, performance evaluation, compensation, selection, grievance resolution, or arbitration.

The lack of cross-fertilization in gender bias research has resulted in isolated clusters of empirical studies within particular research domains. Currently, streams of literature investigating gender bias in evaluation of performance and

evaluation of leaders are developing without reference to the gender bias literature in arbitration and criminology. For example, studies in the context of grievance resolution and arbitration of prominent researchers such as Dalton et al. (1985a, 1985b, 1987) and Bemmels (1988a, 1988b, 1988c) are not cited in research investigating gender bias in leadership or performance evaluation. This is unfortunate, for all such evaluations occur within organizations and are made by decision makers who often have wide discretion. Since all such evaluations have the potential to be tainted by gender stereotypes, an integrative approach to gender bias research is needed.

This study will not only add to the academic literature on gender bias, but also should have important implications for organizations. Most public and private organizations today have policies that seek to protect workers from being unfairly evaluated in different forums. These policies have largely come about due to changing social norms about fairness in the workplace and in response to various equal opportunity laws passed by Congress in the last three decades.

The policy changes in organizations, however, have not always been effective in neutralizing biases against women in the workplace. Top managers, as well as policy makers, can benefit from research results in this area as they devise strategies to tap the talents and energies of the growing number of female workers and managers in the labor force. For

example, if research results indicate that biases against females are much stronger in certain evaluation contexts than others, this would have clear implications for how organizational resources should be allocated to counter such biases.

Findings from research on sex stereotypes over the last several decades have proven to be very useful in demonstrating when biases against women are most likely to take place. The practical impact of these findings became even more apparent in 1989 as the Supreme Court benefitted from this research in making its decision in the Price Waterhouse v. Hopkins case. The present study seeks to continue in this vein and contribute both to the academic literature and to practitioner concerns.

## **CHAPTER 2**

### **LITERATURE REVIEW**

#### **INTRODUCTION**

The literature on the influence of gender on a variety of work-related outcomes such as performance evaluation, managerial and leadership effectiveness, career paths, power in organizations, and grievance outcomes is voluminous. Although the thesis that gender stereotypes bias judgements of decision makers is almost universally accepted, the results of many studies reveal that gender bias is a complex phenomenon. Indeed, research results on gender bias in the past have been inconsistent and are inexplicable in terms of a simple model (Swim et al., 1989; Top, 1991).

This review will focus on two distinctly different streams of literature that have investigated gender bias in different settings and within different frameworks. First, an important Supreme Court decision that addressed the issue of sex stereotyping in the workplace and is influencing current research on gender bias will be summarized. After setting the context for the discussion, findings from the literature on how gender biases may affect the evaluation of individuals or their products will be reviewed. This will be followed by a review of the leadership evaluation literature as it pertains to gender bias. In the second part of the review, findings on

gender effects from the criminology and arbitration literatures will be reviewed and summarized. Implications these findings may have for workplace justice in the context of grievance resolution by managers will be discussed. An attempt will then be made to integrate the results of these streams of literature using models from social psychology.

### GENDER BIAS IN EVALUATION OF PERFORMANCE AND LEADERSHIP

Whether gender stereotypes influence evaluations of female workers and managers is an important question in our society and has been widely debated in scholarly journals for decades. The findings of research in this area have eminently practical as well as legal implications. The next section discusses an important decision by the Supreme Court that has impacted the debate on the role gender biases play in evaluations of women. The High Court's decision brought into sharp focus the important part research in the social sciences can play in influencing policy decisions.

#### The Price Waterhouse Case

In 1989, the landmark Supreme Court decision in Price Waterhouse V. Hopkins spoke to the issue of sex stereotyping at the workplace. The Court ruled that once the plaintiff (Hopkins) had met the burden of proving that sex was a factor in the employment decision to not promote her, the employer

(Price Waterhouse) had the burden of proving that it would have made the same decision even if sex had not been a factor. The case was remanded by the Supreme Court to be decided in compliance with its ruling and instructions on the burden of proof that would be required of the employer. The final decision in the case was rendered by Judge Gessell of the federal district court in May of 1990. Judge Gessell stated plainly, "The firm of Price Waterhouse refused to make Ann Hopkins a partner. Gender-based stereotyping played a role in this decision" (Hopkins v. Price Waterhouse, 1990, p.1). The facts of this historic case are summarized below, paraphrasing Reed (1989, p. 112).

In 1983, Price Waterhouse had 662 partners nationwide, of which only seven were women. In that year there were 88 candidates for partnership. Eighty-seven of the candidates were males and one was female. Ann Hopkins, the single female candidate for partnership, had been responsible for generating more business than any other candidate. Prior to being nominated, she had been a senior manager for four years at the firm's Washington, D.C. office. Despite her successful record as measured by objective criteria, Hopkins was denied partnership on the grounds that she lacked interpersonal skills. Ann Hopkins left Price Waterhouse in 1984 when she was not renominated for partnership. The same year she sued Price Waterhouse on the grounds that she had been rejected as a

partner due to "sexual stereotyping."

During the trial, Price Waterhouse presented evidence indicating that Ann Hopkins was very aggressive, at times harsh, was impatient with staff, and was difficult to work with. Ann Hopkins was able to document that she had been referred to by a partner as "macho" while another had told her that she "overcompensated for being a woman." A third partner had told her to "take a course at charm school." Her use of profanity had been criticized by some partners. Hopkin's supporters felt that her swearing had been objectionable only because she was a female. Her friends in the company advised her to "walk more femininely, dress more femininely, wear makeup, have her hair styled, and wear jewelry." Based on the presented evidence, the district court found Price Waterhouse guilty of violating Title VII of the Civil Rights Act in denying Hopkins a partnership in the firm. After a protracted legal battle involving appeals all the way to the Supreme Court, Ann Hopkins finally won in 1990.

Price Waterhouse v. Hopkins was the first sex discrimination case decided by the Supreme Court in which evidence from social science literature on sex stereotypes played a key role in the final decision. Indeed, the psychological literature on stereotypes was cited even at the earliest stages of the case (Hopkins v. Price Waterhouse, 1985). When the case made its way to the Supreme Court, the



Court extensively discussed the impact of sex stereotyping on employment opportunities of women in organizations. This discussion was informed by the *amicus curiae* brief filed by the American Psychological Association (APA) which reviewed some important literature on sex stereotypes. The very significant decision by the APA to join the case was based on several factors. Most of all, however, the case was important from the perspective of social psychology, and APA had special expertise to address the issue of sex stereotyping in the workplace. The APA especially wanted to inform the Supreme Court that the literature on sex stereotyping, cited by the expert witness (Susan T. Fiske, a noted social psychologist) during the trial, was accepted as valid in the scientific community (Fiske et al., 1991).

The APA brief stated that sex stereotyping can indeed bias the evaluations of women. For example, successful performance on the part of women may be attributed not to talent or skill, but to external factors such as luck or having an easy task. On the other hand, successful achievements of men at work are typically attributed to more enduring internal factors such as ability. Furthermore, the brief argued that, in traditionally male-dominated jobs, the performance of women may be devalued simply because they are women. Clearly, the Court was influenced in its decision by previous research findings on sex stereotyping and gender

bias.

The brief filed by the APA in the Price Waterhouse case has provided an important context for a discussion on gender bias in the workplace and implications this may have for the careers of women. It is well established that women are under-represented as managers and leaders in the upper levels of organizations (Bergman, 1986; Nivea and Gutek, 1981; Powell, 1988, 1990). There is evidence that when the educational level is equivalent, women earn less on average than men (U.S Bureau of Census, 1987). Recent research indicates that even when female managers follow the traditional male model of career advancement and continuously stay in the work force, their salaries increase at a lower rate than those of their male counterparts (Stroh, Brett, & Reily, 1992).

Scholars have further suggested that when female managers engage in leadership behaviors similar to those of male managers, they may be negatively evaluated relative to their male counterparts (Eagly et al., 1992). Indeed, the Supreme Court decision in Price Waterhouse, in agreement with social psychological research on sex stereotypes, stated that the aggressive and forceful behavior of Ann Hopkins in her professional role would have been interpreted as acceptable and looked upon quite favorably had she been a male (for an informed discussion of the case, see Fiske, Bersoff, Borgida, Deaux and Heilman, 1991). The review of literature that

follows will shed some light on the plausibility of this view. The validity of other perspectives that hold that sex stereotypes lead decision makers to differential evaluations of men and women in their various roles in the organization will be carefully examined.

### The Goldberg Study and its Replications

More than two decades ago, in an influential and an often-cited study, Goldberg (1968) demonstrated the possibility of prejudicial evaluations of work done by women even when evaluators themselves were women. Using a simple experimental design, Goldberg gave female undergraduate subjects identical booklets containing six different journal articles in different fields. These fields were either sex-linked (male dominated or female dominated) or sex-neutral. For every one of the six articles, half of the female subjects were told that the author was male and the other half were told that the author was female. This was accomplished by manipulating the names of authors of the journal articles. The results showed that male authors received higher ratings than female authors in most occupational fields. In particular, the difference in ratings was larger in the fields perceived to be masculine.

Goldberg's findings, clearly intriguing, inspired numerous replications of his study with male as well as female

subjects. The results from this stream of literature, however, have not been clear or consistent (see, Walston & O'Leary, 1981, and Basow, 1986, for qualitative reviews). Swim, Borgida, Maruyama, and Myers (1989) have complained that the results of the Goldberg study have often been distorted and exaggerated by scholars. Specifically, they argue that authors in both the scientific and popular literature have misrepresented the strength of Goldberg's original findings.

Swim et al. (1989) attempted to answer the question of whether gender stereotypes bias evaluations by conducting a meta-analysis which included 123 studies from 106 published journal articles. All of the studies used the classic Goldberg (1968) paradigm in that they were true experiments. These studies spanned the time period from 1968 to 1985, but a vast majority of them were published between 1973 and 1981 in North American English language journals.

The meta-analysis by Swim et al. (1989) is the most comprehensive attempt to aggregate results from gender bias studies in different domains. The number of subjects involved in these studies ranged from 20 to 3,261. The total number of subjects in the 123 studies was 21,379. When the three studies having more than a 1000 subjects are excluded, the mean number of subjects per study is 173 and the median is 96. Nine percent of studies used only female subjects, 17% used only male subjects, and 72% involved both males and females.

Seventy-four percent of the studies used college aged students, 21% of the studies involved subjects older than college students and 4% involved subjects younger than college students (Swim et al., 1989).

In about half of the studies, subjects were asked to evaluate only male or female target persons. In the other half, subjects evaluated both male and female target persons. Subjects were generally given evidence of the hypothetical target person's abilities in some form. They were then asked to evaluate the target person, or some work the target person had completed, or both. Evaluations were typically made of written work, art work, behavior, resumes, job applications, or short biographies of people (Swim et al., 1989).

Results of the meta-analysis indicated that the average difference between ratings of men and women is negligible. The authors also found no significant difference between male and female subjects on their ratings of male and female target persons, although the findings for male subjects were much more heterogeneous than for female subjects. There was a greater difference in ratings when the stimulus materials were masculine rather than feminine, with men being rated higher than women. Interestingly, the largest effect sizes were found when the stimulus materials were sex neutral, although even these were relatively small. The authors found that women were rated less favorably than men when less information was given

to the experimental subjects. This is consistent with findings from previous research (Tosi & Einbender, 1985; Locksley, Borgida, Brekke, & Hepburn, 1980; Rasinski, Crocker, & Hastie, 1985) which suggest that the use of stereotypes in making evaluation decisions diminishes when more relevant information is provided about the target persons.

Swim et al. (1989) point out that recent research results indicate the continued existence of gender stereotypes among the general population and the willingness of people to express stereotypes (e.g. Deaux & Lewis, 1983; Martin, 1986). In light of this, they feel it would be a mistake to conclude, based on their meta-analysis, that people in fact do not hold gender stereotypes. Several explanations are offered by Swim et al. (1989) for the meta-analysis results. First, it is possible that Goldberg's paradigm is not sufficient to reveal gender biases among the experimental subjects. Second, since Goldberg is widely cited, it is possible that college students in introductory psychology classes, who generally make up the overwhelming majority of experimental subjects, are sensitized to sex discrimination and gender bias issues. Therefore, their responses may be colored by their desire to appear unbiased (Swim et al., 1989).

The conclusion by Swim et al. (1989) is that there is not enough evidence to support the simple prediction that subjects evaluate men and women differently in the context of the

Goldberg paradigm. However, the authors suggest that gender stereotyping may be a more complex phenomena than envisioned by many scholars. It is pointed out that some set of factors not systematically investigated in the studies included in the meta-analysis may be responsible for triggering gender bias (Swim et al., 1989).

A qualitative review of literature on the evaluation of the scientific, artistic, and literary products of males and females by Top (1991) echoed some of the findings of Swim et al. (1989). Top, like Swim and others, concluded that no consistent overall tendency to evaluate men more favorably than women can be readily ascertained. Therefore, it is likely that gender bias is a more complex phenomenon than first suspected by researchers. According to Top (1991), manifestation of bias against females may depend on characteristics of the stimulus material or person, characteristics of judges, or the situation itself in the context of which the judgment must be made.

### **The Role of Organizational Social Composition**

Since the meta-analysis by Swim et al. (1989) and the qualitative review of Top (1991), several important studies on gender bias have been published. One interesting study (Sackett, Dubois, & Noe, 1991) addressed the question of whether work group representation had an effect on male and

female performance ratings. Their finding was that the social composition of the work group by gender was significantly correlated to performance ratings of men and women. The Sackett et al. (1991) study came about as a reaction to the critique of the brief filed by the APA in the Price Waterhouse case (Fiske et al., 1991; Sackett et al., 1991), in which the APA had suggested that the social composition of work groups in an organization could influence the behavior of its members. It would be useful to mention briefly the contents of the APA brief and the nature of criticism directed at the brief before proceeding further.

The APA, in a review of the literature on sex stereotyping (Gutek and Morasch, 1982; Heilman, 1983; Kanter, 1977; Terborg, 1977), had suggested that gender stereotypes would play a stronger role in influencing evaluations when the number of women in the work setting was comparatively small. Kanter (1977) described a work situation as "skewed" when 15% or less of workers belonged to a minority group, and as "tilted" where between 15% and 35% of the workers belonged to a minority group. The APA brief, relying on this literature, had argued that employment decisions are most likely to be discriminatory in "skewed" situations and that the same process would be present in a "tilted" situation but to a lesser extent. This has been referred to in the literature as the tokenism phenomena. The implication is that when the



percentage of workers belonging to a particular category (by gender or race) is quite small, it exerts a powerful influence on how they are perceived and evaluated.

The APA was subsequently criticized for not making it clear in the brief that the so-called tokenism phenomenon had only been found to exist in laboratory situations where typically college students serve as experimental subjects (Sackett et al., 1991). As a reaction to this critique of the APA brief, Sackett et al. (1991) initiated a study to determine whether the laboratory results indicating the existence of a tokenism phenomena could be replicated in the field.

Sackett et al. (1991) used a data base that had been created by the U.S. Employment Services (USES) for the purpose of examining the validity of General Aptitude Test Battery (GATB). This data base included vital information, such as GATB scores, job performance measures, and demographic measures for more than 36,000 individuals in 174 jobs. Overall, there were 2,876 job-firm combinations in the data base. (Note that the job-firm combination refers to a group of workers having the same job and working for the same firm.) Almost all of the jobs contained in the data base were blue collar and clerical jobs. The data had been collected from 1972 to 1987 (Sackett et al., 1991).

To analyze the difference between male and female

performance ratings, 486 of the 2,876 job-firm combinations were selected. The sample met the following three criteria of selection. First, the job-firm combination had to have some variation on the sex variable. Second, supervisory ratings must have been used to measure performance, and not some test score. Third, each job-firm combination had to contain at least 10 jobs but not more than 100 jobs. The unit of analysis in this study was the job-firm combination. One important part of the analysis focused on measuring the performance of women relative to men (Sackett et al., 1991).

The results of the analysis indicated that the overall difference between men and women across job-firm combinations was very small, with women receiving slightly lower ratings than men. However, performance ratings for women were significantly lower when women constituted less than 20% of the work group. This did not hold to the same extent when men constituted less than 20% of the work force. In other words, the results show that the tokenism pattern for men and women is not symmetrical. Nevertheless, it should be noted that when women made up more than 50% of the work group, they did receive higher performance ratings on average than men (Sackett et al., 1991).

Even when cognitive ability, psychomotor ability, and firm experience were taken into account, the proportion of women in the workplace still accounted for an additional 4% of

the variance in performance ratings. The authors noted that the group composition does not work similarly for race (white-black) as it does for sex. The average difference between black and white ratings was much larger and did not decrease to being negligible as the percentage of black workers increased in the work group (Sackett et al., 1991).

Despite the various limitations in the study, the authors concluded that it offers evidence that the composition of the work force by sex can affect performance ratings of men and women. More specifically, if women constitute less than 20% of the work group, their performance ratings relative to men may be significantly lower. This result from the field supports the laboratory results that have used college students as experimental subjects to test the influence of group composition by sex on performance ratings (Sackett et al., 1991).

### Overvaluation of the Performance of Women

The Sackett et al. (1991) study raises an interesting issue of the overvaluation of the performance of women relative to men when the work group consists of more than 50% females. Although it is widely accepted that, in general, men are evaluated somewhat more favorably than women, there have been several studies indicating more favorable evaluations of the work of women relative to that of men under certain

conditions. For example, Peck (1978) found that female subjects gave high-status women higher ratings on their work than their male counterparts. The same did not hold true for low-status women, for their work was devalued. Along the same lines, Issac (1981) found that women in non-traditional masculine fields received lower evaluations from male judges if the judges had been led to believe that these women had low professional status. However, if the judges thought that women in these masculine fields had high status, then no bias was present in their evaluations.

Overvaluation of women relative to men has also been found in the selection literature (although it is certainly not the rule; see for example, Olian, Schwab & Haberfield (1988) for an insightful meta-analysis). Gender biases in favor of women appear to occur particularly when women apply for jobs which are extremely male sex-typed, and when positive information regarding the competence of the female applicant is available (see for example, Krygar & Shikiar, 1978; Heilman et al., 1988). Heilman et al. (1988) have argued that such overvaluation is consistent with the literature on stereotypes. Researchers have maintained that when the observed behavior of an individual is completely incongruent with stereotypical expectations, it can lead to a sharp reaction in the opposite direction of the one that would normally be expected (Feldman, 1981; Weber & Crocker, 1983).

Evaluators faced with a competent female applying for a traditionally male job have to deal with observations (the documented skills and abilities of the female applicant) which diverge sharply from their stereotypical expectations about females. The extreme discrepancy between observations and expectations can set into motion gender contrast effects resulting in a more favorable evaluation of the female candidate over an equally qualified male candidate (Heilman et al., 1988).

There has been some discussion in the literature of the gender contrast hypothesis (Etaugh & Sanders, 1974) and what has been dubbed "the talking platypus phenomena." "It matters little what the platypus says, the wonder is, it can say anything at all." (Abramson, Goldberg, Greenberg, & Abramson, 1977, p. 123). The idea is that performance on the part of a female in a male domain may be overvalued due to the element of surprise. Heilman et al. (1988) have pointed out that in order for the female to be overvalued, she must be applying for a highly male sex-typed job. In addition, there must be clear documentation of her abilities. Only then will the gender contrast process be set into motion. Thus, it seems that overvaluation of women can be demonstrated under laboratory conditions. However, the conditions under which such overvaluation is likely to occur appear to be quite limited and rather exacting (Heilman et al., 1988).

### Gender and Evaluations of Managers

Much of the focus of the review so far has been on the existence of possible biases in the evaluation of female workers. In order to paint a more complete picture of gender biases in organizational decision making, it is vital to specifically include female managers in this discussion.

The reason for distinguishing between female workers and female managers is simple. In organizations, managers have higher status than workers. It follows that female managers have higher status than female workers. As mentioned in the previous section, researchers have found that gender and status characteristics may interact to produce differential evaluations of men and women (e.g., Peck, 1978; Issac, 1981). Based on this, it can be argued that gender biases in the evaluation of female workers and female managers may be of a differing quality, differing intensity, and may even lead to differing outcomes. In light of this reasoning, it seems important to conduct a finer analysis of gender bias and to distinguish between female workers and female managers in organizations.

An important meta-analysis addressing exclusively the issue of the influence of gender on the evaluation of leaders was conducted recently by Eagly, Makhijani, & Klonsky (1992). The authors looked at studies that used experimental designs where the gender of the leaders was varied holding constant

their other characteristics. This meta-analysis was more specific in its focus than the general meta-analysis by Swim et al. (1989) referred to earlier, which did not include 42 of the studies included in this more recent review of gender bias. This was partly due to the fact that Swim et al. (1989) did not include unpublished manuscripts or dissertations in their meta-analysis (Eagly et al., 1992).

The Eagly et al. (1992) sample included 56 published and unpublished documents that reported the results of 61 studies. The results of the meta-analysis indicated that, across the entire spectrum of studies in the leadership literature, male leaders were only slightly more favorably evaluated over female leaders. The authors noted that the overall mean estimates were very close to the estimates derived by Swim et al. (1989) in their meta-analysis, which consisted of a more general sample of studies in which subjects evaluated male and female performance and behavior. The largest differences between male and female leaders were on general evaluative measures such as competence and satisfaction with the leader (Eagly et al., 1992).

Although the general trend of male leaders being more favorably evaluated was small, under specific conditions the difference between evaluations of male and female leaders was large. For example, leadership style led to a differential in male and female ratings. When leaders behaved autocratically,

subjects gave significantly higher ratings to male leaders than female leaders. On the other hand, when leaders used a democratic and a consensus-building style, subjects rated both male and female leaders equivalently. Also, male subjects showed a stronger tendency to devalue female leaders than did female subjects. Female subjects, on the other hand, did not show any particular gender bias in their evaluations (Eagly et al., 1992).

Eagly et al. (1992) also did not find any evidence that increasing the amount of individual information given to subjects about male and female leaders weakened the bias against female leaders. This is inconsistent with other literature in this area (e.g., Tosi & Einbender, 1985). The authors point out, however, that in all of the studies examined, the subjects had a moderate amount of information about the target persons. The authors suggest that it is possible stereotypical conceptions come into full play only when the perceiver gets very little information about the target person in the experiment.

Overall, the conclusions of Eagly et al. (1992) are that female managers and leaders in organizations may be quite disadvantaged in comparison to male managers and leaders under specific conditions and when adopting certain leadership styles. The authors also point out that biases against women in the work place may be much stronger than those exhibited in



the experimental setting in the laboratory, where the college students may be sensitive to being under scrutiny and may moderate their responses accordingly. Furthermore, even though evaluators in organizations have much more information available to them about the male or female manager to be evaluated, the information may not necessarily dampen the gender bias. Based on Snyder (1984), the authors suggest that an early negative evaluation of a female manager due to limited information may set up an initial unfavorable impression. This may, in turn, prejudice later evaluations, even though much more information may be available about the female manager at that point (Eagly et al., 1992).

In summary, the review of gender bias literature in the performance and leadership domains reveals that gender bias is not a simple phenomenon. Although overall men are only slightly favored over women, under specific conditions these differences can be large and meaningful and may have serious implications for the careers of women. Women are more likely to be rated lower than men on performance as well as managerial ability if the work group composition has a low percentage of women (less than 20%). A woman who violates the gender-role stereotype in her managerial behavior is also likely to be rated lower on leadership ability than her male counterparts under most conditions. Overvaluation of the performance of women may take place if the work group

composition is more than 50% female. Also, under certain conditions, a woman engaging in behavior that clearly violates gender stereotypical conceptions may be overvaluated compared to a similarly qualified man, if there is clear evidence of her competence.

### GENDER BIAS IN DISPENSATION OF JUSTICE

As noted before, in the evaluation of the performance of workers or their managerial and leadership abilities, a slight overall gender bias is found in favor of men over women. However, a somewhat mixed pattern emerges when penalties for individuals have to be decided in response to mistakes committed by them on the job. There has been some research conducted on how gender may influence the outcome of a grievance at the workplace. Since this is a relatively new area of organizational research, many scholars have relied on research results obtained from a branch of sociology known as criminology to formulate hypotheses and offer possible explanations for the outcomes of grievances in organizations.

### Gender Bias Models from Criminology

In the criminology literature, it has long been recognized that the defendant's personal characteristics, such as race and age, can be factors in the treatment received from authorities (Sellin, 1928). Interest in the influence of

gender on case disposition is more recent and appears to have increased dramatically during the 1970s (Nagel and Hagan, 1983). Various research strategies, such as field studies, field experiments and laboratory experiments, have been employed to assess the impact of gender bias on decision making in the judicial system.

There are two major models of judges' decision-making behavior in the criminology literature which are relevant to the discussion on gender bias (Bemmels, 1988a; Moulds, 1980; Nagel & Hagen, 1983; Edwards, 1989; Johnson and Scheuble, 1991; Crew, 1991). Both models assume that male decision makers are affected by gender stereotypes in making their decisions. The first is the chivalry/paternalism thesis. The chivalry concept refers to a protective attitude men are assumed to have toward women. According to this view, men are reluctant to accuse women, inflict harm on them, or punish them for their offenses. The paternalism thesis holds that men view women as defenseless, somewhat like children, who are in need of protection and who should not be held responsible for their actions (Moulds, 1980). Researchers have argued that it is difficult to empirically distinguish between chivalry and paternalism, since they result in similar outcomes (Nagel & Hagen, 1983). Therefore, criminology scholars often refer to the two concepts jointly. The chivalry/paternalism thesis implies that male decision makers will try to protect females

from the criminal justice system and, therefore, will tolerate their offenses more than the offenses of males.

A second competing model found in the criminology literature is known as the evil-woman thesis (Nagel and Hagan, 1983). According to this view, male decision makers will punish women not only for their offenses but also for their inappropriate gender-role behavior. Thus, the same unlawful behavior of males and females will be responded to differently if traditional sex roles are violated (Crew, 1991; Edwards 1989; Johnson & Scheuble, 1991; Nagel & Hagen, 1983). This thesis implies that women will be judged more harshly than men for committing the same offenses and are likely to be given longer sentences after guilt has been established.

Although the results have not always been consistent, many authors, relying on the field studies in criminology, have concluded that women are treated more favorably than men in the criminal justice system. For example, Steffensmeier (1980) gave the following assessment of the situation: "A survey of criminology literature over the past century reveals that the paternalistic thesis is considered virtually an iron law" (p. 340). According to other criminology scholars, women are less likely to be arrested, more likely to have their charges reduced or dismissed altogether, less likely to be convicted, and tend to receive more lenient sentences than male offenders (Moulds, 1980; Nagel and Hagan, 1983). Along

the same lines, Chesney-Lind (1987) stated, "evidence of lenient sentencing of women charged with felonies has been found consistently, particularly when the judge is confronting a decision to incarcerate" (p. 134).

A recent field study (Johnson & Scheuble, 1991) investigating gender bias in the disposition of juvenile court referrals and using a sample size of 36,680 juveniles concluded that there was evidence of gender bias across time, location, offense committed, and previous referral to the court system. The results supported the view that chivalry was a persistent force in court decision making, leading in general to more lenient sentencing and penalties for female juvenile offenders. On the other hand, sex-role traditionalism model which would predict more severe punishment for female juvenile offenders, was on the decline (Johnson & Scheuble, 1991).

Although there is evidence that women receive more lenient sanctions from the court authorities, it has also been pointed out that there are exceptions. For example, crimes such as prostitution, which involve sexuality, often elicit unsympathetic reactions from authorities (Edwards, 1984). One study found that while women were treated more leniently than men for certain felonies, they were dealt with more severely for child abandonment or simple assault (Zingraff & Thompson, 1984). In a conceptual review of research on gender bias in

case disposition from a feminist perspective, Edwards (1989) stated, "The role of law in the regulation of sexual behavior seems far more consistent with a need to control the 'evil woman' than with the 'chivalry' hypothesis" (p.170). Thus, it appears that paternalism/chivalry is not a universal rule in the criminal justice system. However, despite the exceptions, inconsistencies, and the apparent complexity of the issue (see for example, Kruttschnitt, 1982, 1984; Edwards, 1989; Daly, 1989), scholars continue to interpret the empirical literature on gender differences in case dispositions as indicating more lenient sanctions for women than men (Crew, 1991; Dalton et al., 1992).

Whether we can draw inferences from the criminology literature on gender biases to the existence of gender biases in grievance resolution by managers in the workplace is an intriguing question. Only a few scholars (Dalton et al., 1985a, 1985b, 1987, 1992; Bemmels, 1988a, 1988b, 1988c) have drawn direct parallels between the organizational and the criminology literature in the context of gender bias. Research in the management literature investigating the influence of gender on grievance outcomes in the work place therefore is quite limited. The review of literature examining whether managers are influenced by gender when resolving grievances is therefore necessarily brief.

### Gender Biases in Management Evaluations of Grievances

Rosen and Jerdee (1975) used 101 bank managers (73 males and 28 females) at a management training seminar to conduct an experiment to determine the influence of gender on evaluation of a grievance. Their finding was that women who used a polite/pleading approach received a less favorable evaluation on the grievance from managers than women who used a threatening approach. Men who used either approach fared better than women who used the polite/pleading approach. The authors concluded that men had more flexibility than women in choosing the way they wanted to pursue an appeal in response to some perceived inequity at the workplace.

In another experimental study by Larwood, Rand, and Hovanessian (1979), additional support was found for unfavorable treatment of women in the context of workplace grievances. The subjects for this experiment were 104 personnel managers (52 men and 52 women) from federal and state agencies. The results of the study showed that disciplinary action was more likely to be taken against women in traditionally female positions than men in traditionally male positions. The results also indicated that women were more likely to be disciplined if they were in non-traditional positions than men who occupied non-traditional positions. The conclusion of the authors was that mistakes at the job made by career women are more likely to be costly to them than

mistakes made by men.

In a field study by Dalton and Todor (1985a) on workplace justice in which they studied 294 actual grievances at a unionized public utility company, the results were quite different than those in the Rosen and Jerdee (1975) and Larwood et al. (1979) experiments. The researchers found that women consistently received more favorable decisions than men. The study however did not make clear what percentage of these favorable decisions for women were the result of management action and what percentage were due to arbitrator decisions. (Bemmels (1988a), in citing the Dalton and Todor (1985a) study, states that results show managers acted more favorably towards women than men in resolution of grievances. Even though Dalton and Todor (1985a) did not specify the level at which each grievance was resolved in their study, it is quite probable that Bemmels' overall assessment is correct since the majority of grievances are resolved before reaching the arbitration stage). Despite its limitations, the Dalton and Todor (1985a) study is seminal because the authors point out that the results from literature on gender bias in criminology and the results from organizational literature on gender bias appear to give rise to contrasting hypotheses on the effects of gender on decision makers.

In another field study on gender biases in grievance resolution, Dalton and Todor (1985b), using archival records,



analyzed two samples of grievances. One sample involved 310 grievances filed by workers over one year in a western public utility. These employees belonged to a large union local. The second sample involved 222 grievances filed by unionized employees of a different company over a period of one year. Unlike the previous study by Dalton and Todor (1985a), this study showed that the gender of the grievant did not have a significant effect on the outcome of the grievance. However, the results showed that the gender composition of the dyad (union and company representatives) responsible for resolving the dispute was significantly correlated to the outcome of the grievance. For example, when a woman supervisor interacted with a male union representative in handling a work-place dispute, the grievant was significantly less likely to prevail.

In a third field study on the influence of gender on workplace grievances, Dalton, Todor, and Owen (1987) reached similar conclusions. The sample consisted of 673 grievances filed by unionized employees of a public utility over a period of one year. As in earlier studies, archival records of the company were scanned to determine the sex of the grievant, as well as sex composition of the dyad (company supervisor and the union representative) responsible for handling and processing the grievance. Results indicated that although the gender of the grievant was not significant in predicting the

grievance outcome, the sex composition of the dyad responsible for resolving the grievance did significantly correlate with the final outcome. For example, dyads containing female supervisors and male union representatives were three times less likely to result in a favorable outcome for the grievant than dyads consisting of male supervisors and male union representatives.

At this point, there is not enough literature on how managers decide grievances to come to any solid conclusions about the direction of gender bias in management decision making. Whereas the two experimental studies (Rosen and Jerdee, 1975; Larwood et al. 1979) resulted in findings of gender bias against women in the grievance resolution context, the field study by Dalton and Todor (1985a) resulted in just the opposite finding. The two subsequent field studies by Dalton et al. (1985b, 1987) found no significant grievant gender effects. However, in these two studies, the sex composition of the dyad responsible for adjudicating the grievance was found to be significantly correlated with the grievance outcome. Complicating the interpretation of this small stream of literature further is the fact that all the field studies by Dalton et al. were conducted in unionized public utilities using archival data. These results may not be relevant to grievance resolution in a non-union environment where there is no third party, such as the union steward, to

process the grievance.

### Gender Biases in Arbitration

Although only a few studies have investigated gender bias in managerial evaluations and decision making in the grievance resolution context, a stream of literature in industrial relations, building to a large extent on the Dalton and Todor (1985a) study, has addressed the question of gender biases in arbitrator decision making. Arbitrators are third party neutrals hired by union and management to decide grievances that the parties themselves cannot resolve. The decisions of arbitrators are final and binding under most circumstances. Therefore, the question of gender bias in arbitration has been considered important by scholars.

Parallels have been drawn by researchers between the arbitration system found in unionized settings and the criminal justice system (Dalton and Todor, 1985a). In both systems, similar types of decisions have to be made about individuals, and those decisions have the potential to be influenced by various kinds of stereotypes (Bemmels, 1988a, 1988b, 1988c).

Bemmels (1988a, 1988b, 1988c) has suggested that the processes of industrial justice are in many ways similar to those of criminal justice. He has argued that since arbitrators are charged with duties and responsibilities that

are very much like those of judges, they must conduct themselves in a like manner. The comparison between judges' decision making and arbitrator decision making is quite natural and has been discussed before (see Jones, 1980; and Elson, 1980). However, Bemmels is the first researcher to capitalize on the similarities between the two professions and to suggest that the theoretical framework to explain judges' decisions may apply to arbitrator decision making as well. According to Bemmels, the two models found in the criminology literature (the chivalry/paternalism thesis, and the evil woman thesis) to explain gender bias in case disposition by judges should be equally applicable to arbitrators.

However, since theoretical predictions about gender effects on arbitrator decision making are unclear (as the models predict in the opposite direction), scholars have relied on empirical evidence to support one of the two models. Bemmels (1988a, 1988b, 1988c, 1991a, 1991b) has found that male arbitrators are more lenient with female than with male grievants, thus offering support for the chivalry/paternalism model. Recently Caudil and Oswald (1991), using a somewhat different methodology, have reached essentially the same conclusion (see also Oswald, 1991).

There have been other studies reporting no significant gender effects in arbitration (Bigoness & Dubose, 1985; Scott & Shadoan; 1989, Bemmels, 1990a), or gender effects in favor

of men (Rodger & Helburn, 1985). Some of these studies are, however, suspect due to small sample sizes and methodological deficiencies. In the Rodgers and Helburn study, which found that males received preferential treatment from arbitrators when compared with females, only 37 cases were reported in the sample; further, the authors did not indicate what proportion of the grievants were female. Bigoness and Dubose, in their lab experiment, used 80 undergraduate students rather than practicing arbitrators as subjects. Oswald (1991), however, found that the decisions of students differ significantly from the decisions of arbitrators, and therefore students should not be used to infer the nature of arbitrator decision making.

Bemmels (1990a) reviewed ten studies estimating grievant gender effects and concluded that gender effects, although less important than factors such as the nature of the offense, the grievant's prior record, and other mitigating factors, are statistically significant in many studies. Caudil and Oswald (1991) reviewed the findings of some of the same studies as Bemmels (1990a), including the results of their own field experiment, and concluded that female grievants received more favorable treatment at the hands of male arbitrators than did male grievants. Although the results are mixed, the preponderance of evidence in the literature seems to suggest that male arbitrators, who comprise an overwhelming majority of practicing arbitrators, do extend some preferential

treatment to female grievants.

### Implications for Management Evaluations and Decisions

Whether the evidence indicating that male arbitrators extend preferential treatment to females can be used to infer that under certain conditions male managers would extend preferential treatment to females in the grievance resolution context raises an interesting question. In fact, a hypothesis of paternalism effect at the management level is not inconsistent with a paternalism effect at the level of the arbitrator.

Dalton & Todor (1985a) found that women are more likely than men to win their grievances before they reach the arbitration stage due to the favorable treatment accorded them by managers. However, if the chivalry/paternalism factor is at work before the grievance even gets to the final stage, then it stands to reason that grievances of women that do get to arbitration would be weaker than grievances of men (see also, Bemmels, 1988a; 1988b). This is because a female's grievance that has some merit is more likely to be resolved in her favor by a male manager before it reaches arbitration. The same, however, may not hold true for male grievants. Logically, this would imply that women should lose at arbitration more often than men. That, however, is not the case based on empirical evidence.

It is certainly possible that female grievants benefit from favorable treatment by male supervisors and managers responsible for resolving the grievance as well as male arbitrators. This phenomena would be consistent with empirical evidence indicating that arbitrators are only slightly biased in favor of women. The results certainly could not show a high level of bias on the part of arbitrators in favor of females if the advantage the female grievants have due to the paternalism/chivalry of male managers results in only the weak cases making their way to the final step of the grievance.

In other words, if the female grievant benefits from the chivalry/paternalism of male decision makers at many different steps of the grievance procedure, then we can assume that the cases making their way to the arbitrator are the type which would not elicit a strong chivalry/paternalism effect. This line of reasoning is consistent with current empirical evidence on arbitrator decision making. As pointed out in the last section, however, studies investigating the role of gender bias in management decision making in the context of grievance resolution are relatively few and no clear consensus exists as to the direction of gender bias. Therefore, whether managers may favor women over men when resolving grievances remains an unsettled empirical question.

## THEORY AND HYPOTHESES:

### AN INTEGRATIVE AND THEORETICAL ANALYSIS

The thesis that males and females would receive differential treatment from decision makers in the work place can best be grounded in the theories of sex stereotypes found in social psychology. Walter Lippman (1922), an American journalist, distinguished between the "world outside and the pictures in our head." He conceived a stereotype to be a simplified way of looking at the outside reality in order to make it more understandable and manageable than it actually is. Since Lippman, many social psychologists have suggested that stereotypes are generalizations which are rigid and oversimplified, and can lead to biased judgments (Allport, 1954; English and English; 1958; Katz and Braly, 1933; Klineberg, 1951).

Stereotypes come into play in social contexts because they allow us to make inferences without extensive cognitive efforts. This idea was recognized by Tajifel (1969), who postulated that stereotypes are categories used to store vast amounts of information that allow us to bring coherence to our reality. According to this view, biases about groups of people characteristic of stereotypes are the natural result of the limited capacity of the human mind to process information.

Research in psychology over the last several decades has established empirically that stereotyping is a natural process



found in individuals that can lead to discriminatory behavior under various conditions (Fiske et al., 1991). Although the literature on stereotypes focused initially on the influence of race and ethnic characteristics, during the 1960s more scholars started to investigate the existence of sex stereotypes which may lead to biased judgements unfavorable to women. Much of the empirical literature demonstrates that women are often unfairly evaluated on dimensions such as performance and leadership ability because people hold sex stereotypical beliefs and attitudes (see, for example, Eagly et al., 1992; Heilman, Block, Martell, & Simon (1989); and Schein, 1973).

### The Nature of Sex Stereotypes

In some of the earlier work on sex stereotypes, scholars argued (see, for example, Bowman, Wortney, & Greyser, 1965; Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972) that warmth and expressiveness are traits that are identified as typically being characteristic of women, while competence and rationality are traits believed by people to be more associated with men. This has clear implications for whether men or women are more likely to be perceived as having the potential to be successful workers and managers.

Indeed, other research in management (Schein, 1973) suggests that attributes traditionally associated with men

such as self confidence, leadership ability, ambition, and forcefulness are associated with successful managers. Women, on the other hand, are not thought to possess the qualities needed to be successful managers (Basil & Traver, 1972; Bowman, Wortney, & Greyser, 1965). There is evidence that these stereotypical conceptions have not changed much over time (Heilman et al., 1989).

Scholars have suggested, based on empirical evidence, that general stereotypes about women that are shared in the culture are not harmless and that such stereotypes can have negative consequences for the selection (Heilman, 1983; 1988) and the evaluations of women in organizations (Nieva and Gutek, 1980). "... stereotypes promote sex discrimination because based on assumptions about women as a group, inferences are made about an individual's attributes and by large, these stereotyped attributes are antithetical to those presumed necessary for success at traditionally male jobs" (Heilman, 1988, p.99).

The notion that the gender of a person is a diffuse status characteristic and carries with it expectations of certain behaviors constitutes the core of many theoretical explanations for sex-based differences. Perhaps the most potent, and important, explanation of why men and women are reacted to differently in organizations comes from Eagly and her colleagues (1987, 1990, 1991, 1992). The discussion in the

following section is guided by the synthesis offered by Eagly et al. (1992) of previous research to explain gender bias in evaluation of leaders. The theoretical formulation derived in part from this discussion, however, is broader and seeks to explain gender bias as it may relate to managers as well as workers.

### Gender Based Expectations

Eagly's gender-role theory (1987) suggests that people's expectations of behavior in general are colored by their beliefs about what is appropriate or inappropriate behavior for men and women in the society. The notion that such beliefs regarding sex differences are deeply ingrained in people has wide empirical support (Eagly, 1987; Heilman, 1983; Heilman et al. 1989; Schein, 1973). Within organizations, however, the expectations regarding what the appropriate behavior for managers should be come into play (Phillip & Lord, 1982; Eagly & Johnson, 1990; Eagly et al., 1992). These two expectations of behavior may interact in such a way in organizations that expectations regarding appropriate behavior for male managers and female managers may differ significantly (Fiske et al., 1991; Eagly et al., 1992).

Differing expectations for male and female managers may work against women in leadership roles. For example, women are expected to be warm and expressive and possess other nurturing

qualities. Managers, however, are expected to behave in an aggressive, competent, and rational manner. Female managers, therefore, may be caught in a difficult situation. Such conflicting expectations of female managers may lead to a devaluation of women in managerial and leadership roles. Assertive or autocratic behavior of female managers may be perceived much more negatively than similar behavior on the part of male managers (Fiske et al., 1991; Eagly et al., 1992).

The view that female leaders, as well as workers, are reacted to both in terms of their sex and their position in the organization is consistent with the idea of "sex-role spillover" discussed by Gutek and Morasch (1982). Sex-role spillover has been defined as the "carryover into the workplace of gender-based expectations for behavior that are irrelevant or inappropriate to work" (Gutek and Morasch, 1982, p. 55). Gutek and Morasch argue that sex-role spillover occurs for many reasons. First, gender identity is a more basic cognitive category than work role. Therefore, a person views himself or herself (and is viewed by others) as a man or a woman first and a worker second. Second, women may feel more comfortable in the stereotypical female role in the workplace if they believe that by acting in this way they would be better accepted. Third, since historically men have interacted less with women in work roles but more in the roles of

spouses, lovers, parents, and children, it is easy for them to fall in these familiar roles in the workplace (Gutek and Morasch, 1982, p.59).

Eagly et al. (1992) have suggested that the consequences of gender-role spillover are different for men than for women because male managers do not suffer from incompatible role expectations (see for example, Schein, 1973; Heilman et al., 1989). The authors further argue that female managers, when they adopt a leadership style that is democratic and encourages participation, may be able to avoid role conflict at the workplace. While a masculine style of leadership (aggressive, directive, autocratic, etc.) would violate expectations of people regarding a female manager's behavior, a feminine style of leadership would be congruent with her gender-role and, therefore, be more acceptable in the organization. Male managers, on the other hand, since they have legitimacy in leadership roles in organizations, have more freedom to adopt various management styles (Eagly et al., 1992).

Based on the above discussion, it is clear that one of the key constructs in the Eagly et al. (1992) theoretical formulation and analysis of bias in evaluation of leaders is gender-role congruency. It is "defined as the extent to which leaders behave in a manner that is congruent with gender-role expectations" (Eagly et al., 1992, p. 5). The construct of

gender-role congruency is not new and is, in fact, found in many different guises in the literature. For example, Nieva & Gutek (1980) in their comprehensive and insightful review of the evaluation literature pertaining to gender bias stated, "The sex-role congruence explanation for findings of evaluation bias receives support from a number of studies" (p. 272). A scanning of the literature on gender biases reveals that application of the notion of gender-role congruence is not limited to the leadership context. Indeed, the usage of this concept is so natural and ubiquitous that any theoretical framework to interpret the divergence of findings in the gender bias literature must necessarily include it.

### The Role of Organizational Variables

A comprehensive theoretical analysis of gender bias in evaluations can not, however, solely rely on explanations based on gender-roles because people act out their roles as workers or managers in the context of organizations. Organizations differ on many dimensions, and these differences have the potential to influence the gender-role perceptions of its members. It is therefore important to examine some key organizational properties mentioned in the literature as they may pertain to gender bias in evaluations. In this context, two organizational variables have been suggested as relevant to include in an analysis of gender bias in organizations.

These are: the social composition of an organization and the organization culture (Johnson, 1992).

It has been found that if the number of women in a work group or a job category are relatively rare, than this can lead to activation of sex stereotypes on part of decision makers (Fiske et al., 1991). This in turn may bias evaluations of women by their supervisors (Sackett et al., 1991). How the numerical proportions of workers belonging to various social categories (by gender and/or race) can influence perceptions and evaluations regarding them has been discussed extensively by Kanter (1977).

Kanter (1977) identified four types of groups representing different categories of people that can exist in organizations. She defined "uniform" groups to be homogeneous on some external variable such as sex or race. For example, work groups consisting of only males or only females would be considered "uniform." A "skewed" group was defined as having a large number of people of one social type. Kanter suggested that groups that have a ratio of 85:15 in favor of the "dominants" are skewed. In Kanter's terminology, "dominants" refers to people belonging to a particular category that enjoys an overwhelming majority at the work place. Similarly, individuals whose social type (by race or sex) is rare are referred to as "tokens." In practical terms "tokens" constitute less than 15% of the total work group. Groups are

referred to as "tilted" if the "tokens" increase in number to about 35% and become a "minority." Groups are considered "balanced" if all social types are equally represented (Kanter, 1977).

Kanter argued that the behavior of the "tokens" is influenced by their being very visible representatives of their social category. Kanter further suggested that the relative numbers of different social types shape perceptions of the "tokens" and may influence their evaluations in organizations. Kanter's analysis has often been cited to explain discrimination against women in work settings where they are numerically rare in a particular job category. The APA brief in the Price Waterhouse case pointed out that Ann Hopkins was one of the very few women professionals working at Price Waterhouse. Because she was a "token" (in Kanter's terminology), perceptions regarding her were highly likely to be biased by stereotypical conceptions of the mostly male partners.

Kanter's theory of proportions, although insightful, has received mixed reviews. It has found support in laboratory studies (Heilman, 1983; 1984) as well as a field study (Sackett et al., 1991). However, critics have contended that numerical dominance by a particular social type may not be an adequate explanatory factor for discriminatory behavior without taking into account institutional dominance (Izraeli,



1983). For example, one status group may be dominant in a particular field (such as women in child care, or men as managers in manufacturing organizations) without being numerically dominant in a particular work group. Izraeli suggests that groups that dominate a field "determine the rules of interactions for strangers who chance to penetrate the boundaries. Neither their culture nor their power is neutralized by numerical reshuffling" (p. 163).

Kanter's conceptualization that an increase in the numerical representation of a particular status group (based on sex or race) would be advantageous to members of that group has also been questioned. South, Bonjean, Markham, and Corder (1982) found that as the number of females increased in the work groups, they received less encouragement for promotions from their male supervisors. In fact, large minority proportions reduced intergroup contact and, thus, had a negative effect on intergroup relations.

Ridgeway (1988) defines organizational social composition in a slightly different and perhaps more practical way. According to her, it is the representation of members of particular status groups in positions of authority in the organization that affects social interactions in various contexts. Gender may be a salient factor in influencing perceptions when at least 75% of the positions of authority are occupied by members of the opposite sex in an organization

(Ridgeway, 1988). Based on this conceptualization, Johnson (1992) predicted that leader-subordinate interaction may differ significantly between male-dominant and female-dominant organizations. More specifically, female managers may be more likely to stay within their gender-role in male-dominated organizations than in female-dominated organizations. This idea awaits a test in the field.

### Causal Attributions and Gender Bias

Researchers have suggested that attribution of causality for performance may differ between the sexes. This in turn may have consequences for how that performance is interpreted (Nieva & Gutek, 1980). Certainly, similar performances on the parts of men and women, if viewed differently by organizations, may lead to a wide discrepancy in compensation and status over a period of time. That such a gulf may develop between male and female managers having similar qualifications has been documented even in recent studies on gender bias in compensation and promotion (see for example, Cannings & Montmarquette, 1991; Stroh, Brett & Reily, 1992).

The attributional approach to understanding organizational phenomena has proven to be highly useful in many different domains, such as leadership (McElroy, 1982), leader-member relations (Martinko & Gardner, 1987), motivation (Dorfman & Stephan, 1984; Weiner, 1985; Teas & McElroy, 1986),

performance appraisal and feedback (Feldman, 1981; Liden & Mitchell, 1985), organizational conflict (Baron, 1988), and arbitration (Bemmel, 1991a). Attribution theories are likely to play an important role in future organizational research. Such theories are uniquely suited to helping understand why individuals may be evaluated differently due to their gender.

The original framework for explaining how people make attributions in the context of achievement was offered by Bernard Weiner (see, Weiner, Frieze, Kukla, Reed, Rest & Rosenbaum, 1971; Weiner, 1974). Weiner's approach assumes that people have an inherent need to causally understand and explain success and failure (their own as well as that of others). Weiner suggested that people use four specific causes to understand outcomes in achievement oriented situations. These are ability, effort, task difficulty, and luck.

The four causes were classified along two dimensions of causality: locus and stability. Ability, for example, was postulated to be an internal and stable cause of achievement (since it is internal to the person and does not fluctuate over time). Effort was classified as internal but unstable (since it is internal to the person but may vary from time to time). Task difficulty was classified as external but stable (since it is external to the person and the level of difficulty of the task does not change over time). Finally, luck was classified as external and unstable (see Russell,

1982; and Weiner, 1983 for a later criticism of this approach). This conceptualization guided much of the early attributional research directed towards understanding achievement in different situations. It had an important influence on research involving gender differences in leadership and performance evaluation.

Causal attributions for performance are often crucial in employment-related decisions such as hiring, terminations, transfers, or promotions. Many empirical and theoretical studies of attributions (Weiner, 1974; Green & Mitchell, 1979) suggest that if the behavior or the performance of an individual is attributed to causes that are internal to the person and stable (such as ability), then such behavior and performance is considered likely to occur again in the future. This inference can be the basis for a positive employment decision in an organization. However, if the behavior or performance of a person is attributed to external and unstable causes (such as luck), then future performance and behavior cannot be predicted with much confidence. The inference that the performance of an individual is not repeatable in the future because it was due to luck or some other external factor can also have an important influence on employment decisions relating to promotions, retention etc. (Nieva & Gutek, 1980).

Several studies have demonstrated that high performances

of women may be attributed to external causes such as luck or having an easy task (Cash, Gillens, & Burns, 1977; Feather & Simon, 1975), while similar performances on the part of men may be attributed to internal factors such as skills or abilities (Deaux and Emswiller, 1974). Good performance on the part of men has also been interpreted as more indicative of general intelligence even when there was not much difference between the performance ratings of men and women (Deaux & Emswiller, 1974). Since general intelligence is an internal and a stable factor, inferences based on it regarding future performance can be made with more confidence.

Many different types of attribution studies indicate that attributions tend to favor men in the work place. Although the results are at times inconsistent, some studies have found that this bias appears to exist even in the case of ineffective performance. For example, unsuccessful performances by men are more likely to be attributed to external factors such as bad luck, whereas comparative performances by females may be attributed to an internal factor such as lack of ability (Feather & Simon, 1975).

The conceptualization of Weiner et al. (1971; 1974) has undergone considerable development since it was first offered (see Russell, 1982, for an excellent discussion of the measurement problems associated with assessing underlying causal dimensions from specific causal attributions). Much of

the refinement of the original framework has come from Weiner himself (Weiner 1979; 1983; 1985; 1986). In a key addition to the original taxonomy, a new dimension of causality, namely controllability, was added to the other two (locus and stability) proposed earlier. Controllability refers to the degree of control that an individual can exert over a cause. "Causes such as aptitude and or luck are not perceived as subject to volitional influence" (Weiner, 1983, p.531). Due to measurement difficulties (see Russell, McAuley, & Tarico, 1987) and practical considerations (Struthers, Colwill & Perry, 1992), the dimension of controllability is often neglected in the applied literature.

It has also been noted in the literature that people use a variety of explanations for the behavior and performance of others (as well as their own) besides ability, effort, luck, and task difficulty postulated originally by Weiner (Ilgen & Klein, 1988). These other explanations may include but are not limited to mood, personality, interpersonal factors, fatigue, and interest (Elig & Frieze, 1975, 1979; Frieze, 1976; Russo, Kelly & Deacon, 1991). The lesson from this research stream is that specific causal attributions that people make for success and failure may differ depending on the nature of the achievement situation. Therefore, the exclusive attention of researchers on causal attributions involving only four factors, i.e, ability, effort, luck, and task difficulty, may

be misplaced.

While many researchers continue to use Weiner's original framework (see for example, Struthers, Colwill & Perry, 1992), other scholars are focusing on causal explanations that have not yet been explored, such as interpersonal factors (see for example, Russo, Kelly & Deacon, 1991). It is quite possible that attributions regarding males and females differ with regard to not only luck, ability, effort, and task difficulty but also with regard to interpersonal skills, personality, sense of humor, etc. Such attributions are probably endless as they are unique to each situation. Differential attributions of behavior and performance due to the holding of gender stereotypes may shed light on the reasons for differential evaluations of male and female workers and managers.

### The Two-Way Gender Bias

The theoretical explanations for the pro-male biases found in evaluation of performance and leadership abilities can also be used to explain the possibility of a pro-female bias in the context of grievance resolution. Eagly's gender-role theory (1987) allows for the possibility that gender bias may work to the disadvantage or advantage of women in the work setting, depending on whether the behavior exhibited is congruent with role expectations. This idea was expressed by Nivea and Gutek (1980), who also concluded that gender bias

could work in favor of the female worker or to her disadvantage depending on whether her performance was consistent with sex-role expectations.

The implication is that, for women, staying within gender role is crucial to being perceived favorably. To the extent that the workplace behavior of a female worker or manager at a particular job or task contradicts gender stereotypes, her evaluations are likely to be unfavorable compared to similarly situated males. The notion that gender bias can work both ways in different contexts and under different conditions is important and fundamental. It serves as a bridge, reconciling the differing results found in the criminology and grievance resolution literature as well as the performance evaluation and leadership literature pertaining to gender bias.

Two-way gender bias may be a function of attribution of causality for behavior. Such attributions may work in favor of males in the context of evaluation of performance and managerial ability, but may work in favor of females in the context of grievance resolution. According to the paternalism/chivalry thesis, male decision makers, due to their holding sex stereotypical beliefs, view females much like innocent children, who should not be held responsible for their actions and who need to be protected from the harsh consequences of their behavior. This view implies an attribution to external causes for the behavior of lower level



females. According to this perspective, male decision makers (judges, arbitrators) are less likely to hold the female offender responsible for her actions and behavior and, therefore, are more likely to be sympathetic to her in dispensing justice.

Based on the chivalry/paternalism model, it can be argued that male managers, to the extent that they are neutral in the work place dispute (like arbitrators and judges), are more likely to resolve a grievance in favor of a female complainant than a male complainant, all else being equal. The chivalry/paternalism factor should be especially salient when the female is a relatively low-level worker whose work behavior does not violate social norms and expectations pertaining to gender. Even if such a worker is not highly competent, the chivalry/paternalism factor may still come into play as not being highly competent does not violate gender-based expectations for females (see Nivea and Gutek, 1980; Lott, 1985).

Integrating ideas from Eagly's gender-role theory (1987, 1990, 1992) with the paternalism/chivalry thesis, the following general rule can be derived: male managers, to the extent that they are not interested parties in the workplace grievance, would be favorably disposed towards the female grievant as long as the behavior of the female grievant had been within the gender-role parameters. To the extent the

behavior of the female grievant is not congruent with gender-role expectations, it should be expected that the paternalism/chivalry model will lose its predictive power.

It could be argued, based on both gender-role theory and the evil-woman thesis, that a female grievant who has behaved in a manner incongruent with her gender may receive harsher treatment in the grievance resolution context in an organization. It is important to note that gender-role theory can accommodate the competing predictions of both the paternalism/chivalry thesis and the evil-woman thesis, and thus can serve as a theoretical framework for explaining two-way gender bias.

It should be pointed out that many criminology scholars have long recognized that the chivalry/paternalism thesis and the evil-woman thesis are not competing but complementary explanations of the way judges dispose cases (see for example, Nagel & Hagan, 1983; Edwards, 1984, 1989). Recently, Crew (1991) offered a combined model which reflects the view that the legal system is dominated by "patriarchy." "A patriarchal model subsumes both the "evil woman" and the "chivalry" viewpoints in order to predict when each one of these effects may operate" (Crew, 1991, p.61).

Unfortunately organizational scholars such as Bemmels (1988a, 1988b, 1988c), Dalton et al. (1985a, 1985b, 1987, 1992) and, Caudil & Oswald (1991, 1992) present the

"chivalry/paternalism" and the "evil woman" as two competing explanations of judges' decision making and then try to apply it to arbitrator decision making. This approach is rather simplistic and represents a fundamental conceptual error as each effect can operate under different conditions in different cases. It is quite possible for one effect to dominate over the other in some cases, whereas the reverse would be true in other cases. It is not surprising that field studies of arbitrator decisions using archival data, where these two types of effects are not distinguished, find only a slight bias in favor of females.

#### Male v. Female Decision Makers

The paternalism/chivalry thesis offers an explanation for the behavior of male decision makers only. How female managers would make decisions in the context of grievance resolution or evaluations of other managers is less clear. Staines, Tavis, and Jayaratne (1974) introduced the notion of the "queen-bee syndrome" and proposed that females in positions of authority, such as those in supervisory roles, would treat females working for them less favorably than males. Terborg (1977) defines a queen-bee as a woman "who has attained success and status in a man's world and views other women as competitors for her position" (p. 636). The implication of the queen-bee syndrome is that female managers will tend to place higher

expectations on women working for them than on men working for them. Furthermore, since female subordinates or other lower-level female workers may be perceived as potential competitors, females in positions of authority may act to hinder the advancement of other women. This should result in more negative judgements and decisions regarding females but not with regard to males.

While the notion of the queen-bee syndrome has a certain appeal due to its somewhat counter-intuitive prediction about the behavior of females in authority, empirical support for it has not been found in the leadership evaluation literature (Eagly et al., 1992) or the arbitration literature (Bemmels, 1991; Caudil & Oswald, 1991). However, it is probably also true that this idea has not had a fair test in the arbitration literature as only a small percentage of practicing arbitrators are females. Most studies of arbitrators typically involve either no female arbitrators or only a very small sample of them. Research involving gender bias where managers rather than students are used as subjects is so rare that no inferences can be made about how female managers would evaluate grievants or managers based on their gender.

The few empirical results from the arbitration literature which indicate that female arbitrators are not biased in their decision making regardless of the gender of the grievant are consistent with the results from the Eagly et al. (1992) meta-

analysis. Eagly and others found that while male subjects showed a greater tendency to devalue female leaders, female subjects showed no gender bias at all in evaluation of male or female leaders. Swim et al. (1989) in their meta-analysis, however, found little difference between the ratings of male and female subjects of male or female target persons. The authors did note, however, that compared with those of female subjects, findings for male subjects were quite heterogenous across studies.

Overall, it seems that the empirical results on gender bias in performance evaluation, evaluation of leaders and managers, and evaluations of grievants do not show much support for the queen-bee syndrome. If anything, the results from the literature appear to imply that female decision makers may be less biased than male decision makers regardless of the gender of the person to be affected by the decision.

### Research Hypotheses

Based on the above theoretical discussion and a review of the empirical results from the literature on gender bias in evaluation of leaders, performance evaluation, grievance resolution, and criminology, numerous hypotheses can be derived. The present study, however, will focus only on the following.

H1: Male and female managers will differ significantly in

their evaluations of non-aggressive female grievants, with males evaluating them higher than females.

H2: Male and female managers will differ significantly in their evaluations of aggressive female grievants, with males evaluating them less favorably than females.

H3: Male managers will evaluate non-aggressive female grievants more favorably than non-aggressive male grievants.

H4: Male and female managers will differ significantly in their evaluations of autocratic female leaders, with males evaluating them lower than females.

H5: Male managers will evaluate autocratic female leaders less favorably than autocratic male leaders.

## **CHAPTER 3**

### **METHODOLOGY**

#### **INTRODUCTION**

This section discusses the research methodology utilized in the study. Issues pertaining to selection of the research design, choice of the sample, development of measuring instruments, and the quality of measuring instruments will be addressed. Because a pilot study was conducted before the actual study took place, a discussion of the nature of the pilot study and its usefulness is incorporated in this section. First, however, the reasons for the choice of methodology and data collection method will be discussed.

#### **COMPARING FIELD STUDIES AND FIELD EXPERIMENTS**

Both field studies and field experiments have been used to infer gender effects in arbitration, although field studies based on published arbitrator decisions tend to predominate in this literature. In the leadership evaluation and performance evaluation literature as it pertains to gender bias, conducting lab experiments using student subjects is the favored method. The various meta-analyses in the domain of leadership and performance evaluation (i.e., Swim et al., 1989; Eagly et al., 1992) typically aggregate results across studies that have made use of the true experimental design in a

laboratory setting. Both methodologies have their advantages and disadvantages. Either one may be appropriate depending on the goal of the research.

### Advantages and Disadvantages of Field Studies

One advantage of field studies, which have been the most frequent mode of research in the arbitration literature, is that the results of the analysis are based on real cases. For example, published transcripts of arbitrator decisions are based on actual hearings where the arbitrator had face to face contact with all the concerned parties, including the grievant (see, for example, Bemmels, 1988a, 1988b, 1988c). Decisions of managers can be examined using similar methodologies, if an organization has formal grievance procedures, keeps good records, and is willing to make such data available to researchers. The Dalton and Todor (1985a) study serves as a good example of how management decisions in grievance resolution cases can be studied using archival records obtained from the company.

The methodology of field studies based on archival data, however, suffers from certain serious limitations. These weaknesses are particularly troubling when the researcher is interested in knowing whether certain factors cause a decision maker to judge or behave in a particular manner. Causal inferences can not be drawn from archival data.



Furthermore, the cases examined in field studies that are based on analysis of historical data are all different and unique (see, for example, Dalton et al. 1985a, 1985b, 1987). If differences due to the gender of the grievant are found in such studies, it is difficult to conclude whether case outcomes are due to managers holding various types of stereotypes about grievants or whether such outcomes are due to differences in the relative merits of each case. To the extent that male and female grievants have inherently different types of grievances and cases, arbitrator or managerial decisions regarding them will not be comparable.

#### Advantages and Disadvantages of Experimental Methodology

Arguments similar to those made above can be applied to the use of archival data to study management decisions pertaining to evaluations of performance and leadership ability of other managers. However, in the leadership and performance evaluation literature as it pertains to gender bias, experimental methodology has dominated the field from the very beginning (for early examples, see, Goldberg, 1968; Rosen & Jerdee, 1973, etc.).

As Eagly (1992) has pointed out, if the objective of a study is to draw causal links between the gender of the manager and evaluations that manager receives, it can not be done in a natural setting. This is because in real

organizations, the behavior of the male and female managers cannot be made equivalent in order to isolate the effect of their gender on subsequent evaluation decisions. The same line of reasoning applies equally well to investigations focusing on the role gender plays in evaluations of lower-level workers grieving a decision. The type of control needed to draw causal inferences can typically only be achieved using the experimental methodology in a laboratory setting.

The most common method of studying gender effects in leadership and performance evaluations involves presenting respondents with written vignettes describing the behavior or performance of workers or managers (Eagly et al., 1992). The gender of the target person along with other variables of interest are manipulated while everything else is held constant. The dependent variables (such as leadership ability, managerial effectiveness, and performance evaluation) of interest can then be analyzed for variance (see, for example, Rosen & Jerdee, 1973; Bartol & Butterfield, 1976, etc.). While this methodology allows for causal inferences, it is often criticized on the ground that it lacks face validity due to the contrived nature of the experiment. Furthermore, some scholars have suggested that the use of undergraduate students as proxies for actual decision makers is inappropriate (Oswald, 1991). This, however, remains a matter of considerable debate. There is some empirical evidence that at

least in the organization behavior and the human resource management literature, results obtained in laboratory experiments are, in general, consistent with those obtained in the field (see Locke, 1986).

Interestingly, some arbitration scholars have recently turned to the experimental methodology which traditionally has been utilized most heavily in social psychological literature. Bemmels (1991), for example, using written vignettes, employed a field experiment to examine gender effects in arbitration. Caudil and Oswald (1991), also using written scenarios, conducted a field experiment in their study of gender biases in arbitration. It is clear that as the interest in a stream of literature shifts from describing decision maker behavior to explaining its causes, researchers find the experimental methodology to be useful.

It should be noted that field experiments as utilized by Bemmels (1991) and Caudil and Oswald (1991) are different from lab experiments because they are in the field and the subjects used are not undergraduate students but the real decision makers. It has been suggested that actual decision makers, even though they may be subjects in a research study, will use the normal decision processes they have developed in the organizational context (Olian, Schwab & Haberfield, 1988). This line of reasoning implies that results of a field experiment, using managers as subjects rather than

undergraduate students, have more face validity and may be more generalizable to decision makers in organizations.

### Appropriate Design for the Study

This research sought to isolate the influence of gender as one of the determining factors in evaluations of managers as well as lower-level workers grieving a supervisory decision. Because the investigation attempted to examine evaluation and decision-making processes in two different contexts, using gender characteristics of both the decision maker and the person about whom the decision is being made, the experimental design was the most practical research strategy available. Further, due to the advantages associated with using actual decision makers as subjects, a field experiment was decided upon as the appropriate method for this study.

### CHOICE OF DATA COLLECTION METHOD AND SAMPLE

Several data collection strategies for doing field experiments are found in the literature. The most common among these are the mail questionnaire and on-site study. The advantages and disadvantages of various data collection methods were carefully considered with reference to properties of the ideal sample required for a study of this nature. This in turn determined the criteria for selecting the appropriate

method to collect data as well as for identifying an acceptable sample for the study.

Since this study sought to investigate gender bias as a cause of differential evaluations of males and females, it was essential to maintain a high degree of internal validity and rule out all other alternative explanations for differences in evaluative judgements and outcomes. The most effective technique for ruling out a great majority of threats to the internal validity of a study is random assignment (Cook & Campbell, 1979; Pedhazur & Schmelkin, 1991). In the case of this field experiment, it was important that approximately an equal number of subjects be randomly assigned to different conditions of interest.

In light of the above reasoning, the ideal sample for this study had to consist of a relatively homogeneous group of managers, approximately half males and half females, who could be randomly assigned in equal numbers to read different vignettes and answer the questionnaires. Furthermore, this group of subjects had to be cooperative, be willing to get involved in the research study and answer all the questions. (Locke (1986) suggested that subjects' willingness to get involved is a key to making meaningful inferences from the data generated.) Three potential data collection methods were considered. The weaknesses and strengths of each one of these strategies is discussed with reference to the criterion of

internal validity.

### Field Experiment Using the Mail Questionnaire

If it is well designed, the mail questionnaire can have many advantages. It is self administering, which saves both time and money. It can be made anonymous to help increase the response rate, especially if sensitive questions are being asked. Extensive pretesting is important to make sure that ambiguous or complex items have been screened out before mailing out the questionnaires (Isaac & Michael, 1981). Both Bemmels (1991) and Caudil & Oswald (1991) used this method to collect data in their field experiments.

Mailed questionnaires, however, have numerous disadvantages, particularly in the context of experimental methodology. One of the most serious is the low response rate of the selected subjects (see, for example, Bemmels, 1991; Caudil & Oswald, 1991). In a field experiment such as this, if subjects who have been assigned to particular groups drop out in large numbers (i.e, do not read the written vignettes, answer the questionnaires and mail them back), the internal validity of the study can be seriously threatened. Random assignment, although an extremely effective technique for creating equivalent groups, cannot control the threat to internal validity posed by the "mortality" of the subjects. Second, with a mailed questionnaire, there is no assurance

that the instructions or the questions are understood as the researcher intended them to be understood. Therefore, the quality of data may be questionable on those grounds. Third, when the data are collected by mail, it is difficult to insure that the person who answered the questionnaire was the one for whom the questionnaire was intended (Isaac & Michael, 1981).

For these reasons, a field experiment conducted by mail potentially faces serious problems. Some of these limitations can be overcome by first writing letters to subjects (in this case, managers) asking them if they would be interested in participating in a study. Only those managers who express a desire to participate would then be randomly assigned to groups of interest. This would to some extent address the problem of low response rate and subsequent subject mortality after people have been assigned to groups. However, there are no guarantees. The possibility that the cell sizes could differ considerably (to the extent that some cells may not contain any observations) in a field experiment conducted through the mail is quite real. Such an outcome could potentially lead to serious problems in data analysis and interpretation.

#### On-Site Study in an Organization

An on-site study in an organization offers many advantages for a field experiment. First is the issue of

control. Since the researcher is present at the site, he/she can make sure that the experimental procedures such as random assignment are adequately being operationalized. Furthermore, instructions to the subjects can be clarified if there is any ambiguity. Due to the presence of the researcher, subjects at the site are likely to become more involved in the study, thereby generating higher quality data.

Since the on-site study takes place in one organization, it can be assumed that the influence of organizational variables (such as social composition and organization culture) is constant for all subjects. Therefore, the researcher need not worry about those sources of variance diluting the responses of the subjects. Probably, the most important advantage of an on-site study is that the response rate can be expected to be close to 100%. The problem of subject mortality resulting in unequal group sizes is virtually nonexistent. This results in the neutralization of a very serious threat to internal validity.

Despite its numerous advantages, an on-site study is typically criticized by other scholars as lacking in external validity. It is often pointed out that results found in such studies are local to the site and may not be generalizable to other organizations. Although that criticism is valid, it should be recognized that this study constitutes a building block in which the focus is explicitly on internal validity.



This research, of course, does not answer all questions, and future studies at different sites and with different samples would be necessary to answer the questions about generalizability.

### On-Site Study at a Management Training Center

Some researchers are able to get access to subjects while the subjects are attending a management training seminar or are going through an assessment center (see, for example, Rosen & Jerdee, 1973). If an experiment can be conducted with such subjects, the results can be promising. In such a setting, there is perhaps less concern on the part of top management that valuable time is being lost by the workers on the job. Furthermore, the researcher can make sure that experimental procedures are properly operationalized to ensure internal validity. Typically in management training seminars, participants tend to be enthusiastic and get involved in the research study. This can result in collection of high quality data, which can lead to meaningful inferences.

An on-site study at a management training center is subject to similar criticisms as an on-site study in an organization with reference to the issue of external validity. Furthermore, if the data at the training center are collected from subjects who are from different organizations and/or who vary in rank (and therefore are not a homogeneous group), the

results may be confounded by organizational variables (such as organizational culture, social composition of the organization, and status or position of the subject in the organization).

#### Method of Data Collection Chosen

Of the three data collection strategies outlined for conducting a field experiment, the mail questionnaire is the weakest because subject mortality (non-response) can pose serious problems for the internal validity of the study. Therefore it was considered unsuitable for this study. Of the other two strategies, a field experiment in an organization with the support of top management was deemed as the most appropriate and attractive for reasons already mentioned.

#### DEVELOPMENT OF MEASURES AND PRE-TESTING

Two general scenarios (case studies) were constructed to be read by the experimental subjects. These vignettes operationalized the relevant constructs of interest having to do with gender of the target person, gender-role congruence, and the context of the decision. The two cases were completely independent in terms of the facts, the context, and the target persons.

One scenario involves an average performer in a lower-level position who has been fired due to a mistake made at the

work place. The worker then grieves the decision of the department manager. The gender of the grievant and gender-role behavior are manipulated in this scenario. In half of the scenarios the grievant is a female, whereas in the other half the grievant is a male. In half of the cases the grievant adopts an assertive and aggressive stance in interacting with the supervisor who accuses him/her of poor performance. In the other half of the cases the grievant adopts a non-aggressive and polite approach in his/her interaction with the supervisor in similar situations. Everything else in the vignettes besides the name and the pronouns is identical. Detailed background and information about the performance of the worker as well as his/her interaction with the supervisor was given. The original vignette with the grievant as the target person was about three single-spaced typewritten pages long.

Preliminary subjects (three graduate students and a secretary) were asked to read the vignette and then answer 31 questions on a 10 point Likert-type scale pertaining to the performance, ability, personality, effort, luck, and task difficulty of the target person. The respondents were then directed to put themselves in the position of an ombudsperson and select among the most appropriate of six options for the grievant reflecting what their decision would be in the situation. They could either reinstate the grievant fully, uphold the discharge, or select from four other alternatives

falling between the two extremes.

The second scenario involves a previously successful marketing manager who has been made the acting general manager of a large health club. The health club is portrayed as facing problems of slow membership growth, which the acting general manager is attempting to remedy without much initial success. This scenario was also originally about three single-space typed pages. The two manipulations performed in this vignette also pertain to the gender of the target person and the gender-role behavior of the target person. First, the manager is portrayed as a female in half of the scenarios and as a male in the other half of the scenarios. Second, the management style is varied, with managers adopting an autocratic, directive leadership style in half of the vignettes and a democratic, participative leadership style of management in the other half.

Again preliminary subjects were asked to read the scenario and answer 31 items on a 10 point Likert-type scale pertaining to the ability, personality, effort, task difficulty, performance, motivation, luck, and environment of the target person in the scenario. Finally, they were directed to put themselves in the position of a top manager and select the most appropriate of six actions regarding the target person in the scenario. The first action involved making the acting general manager a permanent general manager with an

increase in salary. The sixth option involved firing the acting general manager. The other four options fell somewhere between these two extremes.

The three graduate students and the secretary were then interviewed to find out what they thought about the readability of the cases and the items in the questionnaires following the cases. Based on their comments the typing errors were corrected, the cases were lengthened slightly, and more items were added.

The response of the preliminary subjects was generally positive in that they found both the cases interesting to read and the cases held their attention. The items in the questionnaire were perceived to have obvious face validity as they were all clearly related to the scenarios. None of the these individuals was able to guess the nature of the hypotheses from the cases.

After modifications, the cases were given to four undergraduate students taking a labor relations class in the department of management. These subjects were extensively questioned to find out whether the scenarios were clear, easy to read, and had an element of realism. They were also asked about the clarity of items as well as the instructions given regarding the use of the scale. There was general agreement that the written vignettes realistically and clearly portrayed the two work situations. These subjects indicated that to read

one of the vignettes and answer all the questions took about twenty minutes.

Based on the suggestions made in these interviews, some more modifications were made in the vignettes as well as the instrument accompanying it. For example, the length of the Likert scale was reconsidered. After a review of the literature on scaling as well as on measurement of attributions, a 9-point scale was adopted instead of the 10-point scale. A 9-point scale is the most common one used for measuring attributions and evaluations (see, for example, Elig & Frieze, 1979; Russell, 1982; Dobbins, 1985; Russell, McAuley, & Tarico, 1987). Having such a scale, therefore, allows better comparability with other studies on related topics. Also, based on the suggestions made, it was decided that a visual display of the scale would facilitate the accurate completion of the questionnaire accompanying the cases.

The number of items in both the questionnaires that respondents had to answer after they read the scenario was increased to 51 to obtain more information. After a review of the scenarios and the accompanying instruments, a member of the dissertation committee suggested that a 7-point decision scale be adopted for item 51 instead of a 6-point scale. This suggestion was incorporated in the final instrument which then consisted of 50 regular items on a 9 point scale and a final

item (item 51) where the respondents had to pick one of the 7 decision options given to them. Another committee member's suggestion, that both the scenarios have clearly male and clearly female names, was adopted as well. Thus, in the case of the manager, the name of the male target person was changed to Ted and the name of the female target person was changed to Kathy. In the case of the grievant, the name of the male target person remained Joe, but name of the female target person was changed to Jane.

#### Summary Description

Two fundamentally different cases of approximately three and a-half single-spaced typed pages were written. One case involved a manager as a target person and the other one involved a grievant. Because of the two experimental manipulations pertaining to gender and gender-role behavior, four versions of the vignettes focus on the manager and the other four focus on the grievant as a target person. All of the manager cases are identical except for the manipulations. All of the grievant cases are identical as well except for manipulations. Both the manager cases and the grievant cases are independent of each other with regards to the facts of the case as well as the context. Therefore, responding to one case in a particular way should have no substantive bearing on how the subjects respond to a subsequent case.

Specifically, the key actors (target persons) portrayed in the scenarios are as follows.

1. Autocratic Female Manager
2. Democratic Female Manager
3. Autocratic Male Manager
4. Democratic Male Manager
5. Aggressive Female Grievant
6. Non-aggressive Female Grievant
7. Aggressive Male Grievant
8. Non-aggressive Male Grievant

All of these cases were followed by a questionnaire containing 51 items. Although the manager and the grievant questionnaires were similar in format, length, and even in wording of many items, they were fundamentally different as they addressed different contexts, different factual situations and different target persons.

### Theoretical Justification for the Scale

The written vignettes and the accompanying measures were constructed specifically for this project. Although they are unique to the situation, they have been "sensibly" derived (using Nunnally's (1978) terminology) from the literature. The items in the questionnaires attempt to tap all relevant evaluations regarding various aspects of the performance and behavior of the target person in the work context. Both of the



questionnaires contained reversed scale items distributed randomly throughout to serve as "cognitive bumps" so that the subjects would not fall into a response set.

In writing the items, recommendations of Babbie (1990; 1992), Comrey (1988), Devellis (1991), and Judd, Smith & Kidder (1991) regarding scale development were followed. At different stages both the vignettes and the accompanying items were reviewed by graduate and undergraduate students (who served as subjects in the preliminary pre-testing) and were modified several times. The cases and the measures were further examined by the faculty on the committee. There was general agreement that the items were face valid, easy to read and understand, and explicitly related to the evaluation of the target person in the scenario. The measures clearly represented an adequate sampling of the domain of interest, which in this case involved evaluating the target person in the case. In this context, the measures may usefully be conceptualized as being content valid.

The theoretical support for the measures contained in the questionnaires resides in a stream of attribution theory directed at understanding how people interpret events or actions and behavior of other individuals by assigning specific causes to them. For example, it has been suggested in the attribution literature that there are four major causes that are used to explain the performance (failure or success)

of individuals (Weiner, Frieze, Kukla, Reed, Rest & Rosenbaum, 1971; Weiner, 1974). These are: ability, effort, luck, and task difficulty. These causes are perceived to be either internal (ability or effort) or external (task difficulty or luck) to the individual being judged. The same four causes can also be classified as being either stable (ability or task difficulty) or unstable (effort or luck). This framework of how people perceive the causes of failure and success dominated research in attribution throughout the 1970s.

Eventually the exclusive focus on only four causes (ability, effort, luck, and task difficulty) as explanations for the performance of individuals came to be criticized (see Ilgen & Klein, 1988). It became evident that people attribute behavior to factors as diverse as mood, personality, fatigue, interest, and interpersonal skills, depending on what is relevant to the situation (Elig & Frieze, 1975, 1979; Frieze, 1976; Russo, Kelly & Deacon, 1991; Kent & Martinko, 1992).

Elig and Frieze (1979) made a significant contribution when they examined three different methods of measuring causal explanations. They compared the use of importance ratings (where subjects were asked to rate the importance of specific causal attributions on a 9-point scale), percentage of causality measures (subjects indicated how much each cause contributed to success and failure), and open-ended attributions (subjects simply stated the reasons for success

or failure). Their findings indicated the overall superiority of structured measures in terms of reliability and validity in assessing causal attributions. The subjects disliked the percentage method the most. Elig and Frieze (1979) recommended the use of open-ended procedures (despite its weaknesses with regards to psychometric properties) in the pre-testing stage of research to make sure that pertinent causal factors are included in the structured measures (see also Weiner, 1983).

Many scholars following the Elig and Frieze (1979) approach have tried to assess the underlying strength of causal dimensions (locus of control, stability) by adding scores from scales measuring specific causal factors (such as ability, effort, etc.). This approach, although quite intuitive, has been questioned since it assumes that the researcher and the subject in the study view the causal explanations in the same way. For example, although Weiner (1974) originally conceived ability to be an internal and stable factor, it may be viewed by many individuals as something that can be changed through training or experience. For such an individual, ability would be an internal but not a stable factor. Therefore the responses of such a subject with regards to specific causal explanations could not be collapsed into causal dimensions without doing serious injustice to the measurement process. Weiner's original framework does not address such issues.

The conceptualization of Weiner et al. (1971, 1974) has been refined further (see, for example, Weiner, 1979, 1983, 1985, 1986; Russell, 1982; Russell, McAuley, & Tarico, 1987). One of the key empirical findings of this later research is that attributions can be classified along the dimension of controllability (control over the cause v. lack of control over the cause) as well as the original two identified earlier (Weiner, 1974).

Although additional dimensions along which attributions can vary have been proposed in the literature (see, for example, Abramson, Seligman, & Teasdale, 1978; Weiner, 1979; 1985a), only locus of causality, stability, and controllability have found wide empirical support. (see Weiner, 1983; 1986). Russell (1982) developed the causal dimension scale to measure these three aspects of the way people make attributions. The purpose of the causal dimensional scale is to identify a causal structure which may underlie specific causal explanations of behavior and performance such as ability, effort, luck, task difficulty, personality, mood, fatigue, interest, interpersonal skills, etc.

The causal dimension scale has been used in research on attributions with some success, and there is evidence for its reliability and validity (Russell, 1982; Russell et al. 1987). One concern that has emerged, however, is that in many studies

a high correlation has been found between controllability and locus of causality, raising questions about discriminant validity of the scales. Furthermore the controllability measure is generally less reliable than the locus of causality and stability measures (Kent, 1991; Russell et al., 1987). Russell et al. (1987) have pointed out that the correlation between these two dimensions may vary from situation to situation. They have suggested addition of extra items to the controllability sub-scale to increase its reliability. For practical reasons, many scholars continue to use the original taxonomy offered by Weiner et al. (1971, 1974) and focus only on dimensions of locus of causality and stability in interpreting research results (see, for example, Struthers, Colwill & Perr, 1992).

The continuing difficulty with measuring causal attributions was noted by Ilgen and Klein (1988) in their general review of the organization behavior field. "Attributional research in OB has two shortcomings. First there is a tendency to take as given the four-fold model of attribution sources (i.e. performance is attributed to ability, effort, task difficulty, or luck)....A second problem is the measurement of attributions...The dubious psychometric quality of attribution measures may account for the lack of strong and consistent findings in some of the research." (p. 341). Although Ilgen and Klein pointed out the weaknesses in

attribution research as it has been conducted, they failed to appreciate the recent works of Russell et al. (1982, 1986, 1987) that have provided important insights into the measurement of causal dimensions. Russell and McAuley (1986), for example, have pointed out that both causal dimensions and causal attributions may have joint and independent effects on affective reactions to success and failures. This result appears to suggest that there is a need to include items tapping specific causal attributions as well as the three causal dimensions theorized by Weiner (1979, 1983, 1985) in developing and testing attribution theories.

Because in this study the interest was not in developing and testing general theories about attribution but in the nature of two-way gender bias, the use of the causal dimension scale seemed contrived and inappropriate. Furthermore, since the causal dimension scale is a self-perception measure, it must be altered to be used in research where perceptions about others must be assessed. Altering the causal dimension scale to fit unique situations may, however, change its psychometric qualities.

Due to many unresolved measurement problems in attribution research as alluded to earlier, a common sense approach was adopted for the purposes of this study. Although many items were included that may help to reveal underlying causal structures, that was not one of the goals of this

study. Therefore, the main focus in the questionnaire is on tapping evaluation judgments and specific causal attributions. However, following the advice of Ilgen and Klein (1988), the causal attributions have not been limited to the original four (ability, effort, task difficulty, luck) advocated by Weiner (1974). Furthermore, the suggestion of Elig and Frieze (1979) that open ended questions be included in the pilot study (to make sure that all the pertinent causal factors are included in the structured measures) was adopted. It was reasoned that if gender differences were found in how specific causal attributions are made by managers, it would be useful to rigorously assess underlying causal dimensions in future research.

### Choice and Construction of the Dependent Variables

The purpose of this research was to determine if managers (the research subjects) would vary in their evaluations and decisions depending on different factors pertaining to gender. The dependent variables chosen for analysis, therefore, had to accurately reflect the variance in the judgements and decisions of managers due to the experimental manipulations.

One of the main dependent variables of interest expected to accurately represent the underlying attitude of the manager toward the target person was the manager's perception of the target person's performance. Therefore, the 50 item scale

contained numerous measures tapping how the research subjects judged the performance of the target persons (in both of the written cases) and generally evaluated them with regard to other pertinent factors mentioned in the previous section. The second dependent variable measured the favorability of the research subject's decision pertaining to the target person. This was captured by a single item (the last item on the scale, no. 51) that asked the respondents to choose one of the seven possible actions they felt was appropriate in the situation.

The Likert scale used in the questionnaires tapped many other evaluation judgments of research subjects that were expected to correlate with their perception of the performance of the target person. For example, it was expected that measures of perception of performance could correlate strongly with measures of perception of ability as these two scales may be tapping the same underlying perception of the subjects with regards to the target person. In that case the two sub-scales would be combined to construct one measure. The decision to do an extensive pilot study was based to a large extent on the need for determining the psychometric properties of the various sub-scales contained in the 50-item scale. Based on the results of the pilot study, modifications were expected to be made in how the dependent variable measuring the evaluation of the target person in both the scenarios was to be constructed.



The following preliminary dependent variables were conceptualized to tap the perceptions and attitudes of the research subject towards the target person in the written vignette. The initial names of dependent variables are given in bold. The first two pertain to the scenario with the manager as the target person and the next two pertain to the scenario with the grievant as the target person.

1. Evaluation of manager's performance - **EMP**
2. Decision regarding manager's future - **DRM.**
3. Evaluation of grievant's performance - **EGP.**
4. Decision regarding grievant's future - **DRG.**

It was expected that EMP and DRM would correlate highly with each other as would EGP and DRG. However, because of the possibility of a discrepancy between the evaluation of the target person in the scenario and the decision regarding that person, it makes sense to at least initially conceptualize these variables (EMP & DRM in the management case and EGP & DRG in the grievant case) as revealing different aspects of the operation of gender bias.

Note that all the information solicited by the questionnaire following the two cases was not necessary to test the hypotheses. The hypotheses only address the issue of how the managerial judgements and decisions may differ depending on various factors pertaining to gender and gender-role behavior. However, all the information collected by the

questionnaires was considered to be either theoretically relevant or empirically useful for addressing the issues at hand as well as conducting future research. All of the scenarios and the scales are given in the appendix.

### PILOT STUDY

Before undertaking a major research project, it is almost always advisable to conduct a pilot study (Babbie, 1992). The pilot study can serve many useful functions. It can allow the researcher to gauge the effect size (for example, with regards to a particular manipulation) in the actual population, if that information is not available from the literature. This allows the accurate determination of the sample size necessary to achieve certain statistical power.

Even if the size of the sample in the actual study cannot be modified due to practical constraints, the pilot study still can serve many useful purposes. It can give a good indication of how well the procedures for data collection are working and if changes need to be made prior to the start of the actual study. It can serve as a complete rehearsal for the various steps in the research process, from data collection and data coding to data analysis. By going through every step first in the pilot study, the researcher can avoid any potential pitfalls in the actual study.

### Reliability Issues

The pilot study can also give some insight into the nature of the dependent variables and how they should be constructed. Furthermore, the pilot study can give a clear indication of the reliability of the measures that will be used. Having measures with high reliabilities leads to an increase in statistical power (see Devellis, 1991; Lipsey, 1990) and is important to making meaningful inferences about the results of a study.

The concept of reliability is central to the field of measurement (for discussions on the topic, see Devellis, 1991; Nunnally, 1978; Kerlinger, 1986; and Pedhazur & Schmelkin, 1991). There are two types of errors that can hinder the process of measurement. These are systematic and unsystematic errors. Unsystematic errors are random errors that vary in an unpredictable fashion. The more random error there is in a measure, the lower the reliability of that measure will be.

Although there are many approaches to measuring reliability, high internal consistency of a scale is often interpreted to mean that the scale is free from random error and that the items included in the scale are homogeneous. Pedhazur and Schmelkin (1991) suggest that the many different theoretical orientations to studying the relations of items in a scale (internal consistency) and the various analytic approaches all give estimates essentially equivalent to the

results given by a measure of reliability known as Cronbach's alpha (p. 92). A high Cronbach's alpha for a scale indicates that the measure is highly reliable while a low estimate indicates just the opposite.

Although scholars differ somewhat on this issue (see Pedhazur & Schmelkin, 1991, p. 109), typically in social science research, alphas of over .70 are considered acceptable, alphas of over .80 are considered good, and alphas of .90 or above are considered excellent (see the discussions in Devellis, 1991; and Pedhazur & Schmelkin, 1991). Nunnally (1978) does suggest that although low reliability estimates are acceptable in early stages of research, higher reliabilities are necessary if measures are to be used for determining differences between groups. The impact of unreliable measures on the results of the study can be "devastating" (Pedhazur & Schmelkin, 1991).

Because of the clear importance of having highly reliable measures, it is essential to analyze pilot study results and then modify the scale, if necessary, to improve its psychometric qualities. For example, if the reliability of a particular measure of interest in testing the hypotheses is too low, more items can be added to the scale before the actual study.

The pre-testing also allows checking for the item response bias. Different versions of the scale in which the

items are presented in a different order can be used in the pilot study to explore that possibility. Furthermore, in some instances, a pilot study which uses measures that have been soundly derived from the literature, and therefore are content valid based on expert judgement, can serve as a preliminary test of hypotheses. This is particularly true if the subjects in the pilot study are similar to the population of interest under investigation.

In the present study, however, since the sample (of students) in the pilot test was expected to differ greatly from the sample (of managers) in the actual study, the focus was not on preliminary testing of the hypotheses. Instead the data of the pilot study was analyzed to find out three things: whether the manipulations involving gender and gender role were working; whether the two carefully constructed 50 item scales (one for the case of the grievant and one for the case of the manager) based on the literature, did indeed lead to highly reliable measures for the evaluations of the target persons; the extent to which the ordering of the items and the sequence ordering of the cases affected the responses of the subjects.

### The Sample

The sample for this study consisted of 400 undergraduate students at Virginia Tech who were taking upper-level classes

in the College of Business either in the fall or spring of the 1992-93 academic year. Ninety-five percent of the students were graduating seniors in their final year of school. The other five percent of the students came from an undergraduate class in Organization Behavior also being taught in the College of Business. Overall, 55% of the students were males and 45% were females.

Prior to initiating the study, permission from the human subjects committee was sought to use student subjects. Because of the non-sensitive nature of the experiment, the study was exempted from the oversight of the committee. After the exemption was granted, individuals scheduled to teach Business Policy in the spring semester or those who were already teaching Business Policy at the time were contacted with a request to allow class time for conducting of the experiment.

Several professors, as well as instructors, agreed to the use of their students. Some faculty allowed half an hour for the study whereas others permitted the full class time of an hour and fifteen minutes. In classes where the whole time was allocated to the study, the students received both the case involving the manager and the case involving the grievant as a target person. In other classes, where only 25-30 minutes were allocated to the study, the students received only one case, either the case involving the manager or the case involving the grievant as a target person.

### Manipulations

In order to test the operational hypotheses, two manipulations involving the gender and gender-role behavior of the target person were required in each of the scenarios. These variables represent constructs thought to be important in gender bias research and were operationalized in the written cases. In addition, based on the literature, sex of the research subjects was conceptualized as an explanatory variable. The names of the independent variables are given below.

1. Gender of the target person to be evaluated - **TARGETSEX.**
2. Sex of the research subjects - **RATERSEX.**
3. Gender-role congruence of the target person - **GEN-ROLE.**

Based on methodological considerations, manipulations involving the sequence in which the cases were presented to the respondents and the order in which the items appeared on the questionnaires were introduced in the pilot study. These manipulations find extensive support in the scaling literature on context effects (see, for example, Babbie, 1992; Kerlinger, 1986; Nunnally, 1978; and Schuman & Presser, 1981).

It has been suggested that general questions are more likely to be susceptible to order effects than specific questions (Tourangeau & Rasinski, 1988; although, for an

example of order effects on specific questions, see, Schuman & Presser, 1981). Judd, Smith & Kidder (1991), writing on the subject, candidly state "Our ability to predict what questions may be influenced by order effects is not complete but common sense must be applied in questionnaire construction to avoid problems" (p. 247).

Due to the somewhat inconclusive findings in the literature, the pilot study tested for both the sequence effect of the cases as well as the order effects of the questions in the cases. This decision is consistent with the recommendation of Babbie (1992), who suggested that a pilot study be used to become sensitive to the nature and direction of question order effects. Since the questionnaires contained 10 general items and 40 specific items, the most powerful manipulation to discover order effects involved putting all general items last in one form of the questionnaire and putting all general items first in the second form. The names of the manipulations to test for context effects are as follows:

4. The sequence in which the case was read - **SEQ-ORDER**.
5. The order in which the items appeared - **QUES-ORDER**.

### Design and Procedures

The pilot study employed the experimental design. Different versions of the written vignettes described in the



earlier sections were randomly assigned to students at the beginning of each class. In those classes where instructors allowed enough time for both the grievant and the manager cases, the sequence in which the cases were presented to the subjects was recorded. Furthermore, to check for item response bias, the order in which the items appeared was varied in about one third of all the cases in both scenarios. In one form of the questionnaire 40 specific evaluative items relating to the target person were put first, followed by 10 general evaluative items. In the second form of the questionnaire, the 10 general items appeared at the beginning, followed by the 40 specific items. Item 51, which asked the respondents to pick the most appropriate decision out of the 7 options given, remained in the same place in both forms. The question order was recorded and coded onto the opscan forms.

At the start of each class session, students were informed by the instructor of the class that they would have the opportunity to participate in a study about management evaluation and decision making. The instructors typically endorsed the project in a brief introduction and encouraged the students to participate. After the introduction had taken place, the researcher communicated to the students that the study involved reading a case describing a work situation and then evaluating the key person in the case. The students were told that although their participation was voluntary, if they

did choose to participate, they would find the case to be both interesting and challenging as a management decision making case.

After the cases had been randomly distributed, instructions were read out loud to the subjects and visual scales were drawn on the blackboard to facilitate responding to the items. The subjects were instructed to read the scenarios and fill out the responses to items on the accompanying questionnaires using opscan forms. The students were further instructed to fill in their social security number, their name, the name of the class, time the class met, as well as the date. They were also asked to code in their gender in one of the columns of the opscan. Giving all this information was considered quite natural by the students as majority of instructors gave extra credit to the students for participating in the experiment. Therefore, it was important to be able to identify which students participated in the study.

The subjects were requested, if they had time remaining, to write on the last page of the questionnaire the reasons for their selecting a particular decision regarding the target person. The subjects were encouraged to ask questions if they did not understand a particular item or something else relating to the case. Throughout the time that the students were reading the case and filling out the instrument, the

researcher remained present to observe the subjects as well as answer any questions.

The researcher ensured that the students had correctly coded in their gender on the opscan form and that the opscan form and the case remained together when the students handed them back. The opscan form and the case were separated only after coding the particular version of the case to which the student had responded. In almost all the classes, after the students had completed the cases, at least a few minutes were spent asking about the readability and interest level of the case and the ease with which the students had been able to respond to the items on the different scales.

Information about every class session was carefully documented immediately afterwards. This information included the name of the instructor and whether he or she gave extra credit to students for participating in the study. Further, the documentation included the date of the experiment, the number of students participating, whether the case given involved the manager or the grievant as a target person (if only one case was given in the class), the sequence in which the case was given (if both the manager and the grievant cases were given), the question order of the questionnaire following the case, and the impression of the researcher regarding the level of student involvement in doing the case studies. The responses of the students and any unusual comments in the

discussion afterwards with regard to the scales were noted and recorded.

Although most instructors gave students extra credit as an added incentive to participate in the study, there were two exceptions. One instructor substituted the manager and grievant cases in his classes in place of the regular cases assigned for that day. Another instructor simply requested the students to participate, and, since the study was taking place during the first half hour of class time, everyone did. In all the classes, students appeared highly motivated and involved and took great care to read the cases and answer the questions at the end. The overall participation rate was virtually 100% with only one student turning in an incomplete opscan and suggesting that he preferred not to complete it.

#### The Nature and Quality of Pilot Study Data

Due to a high level of student involvement and response, the quality of data produced by the pilot study must be considered excellent for making useful psychometric and substantive inferences. Nevertheless, it is important to note the unique context in which the data was collected and its implications for the resulting structure of the data.

The pilot study was conducted in many different classes of varying sizes. Furthermore, there were major differences in the amount of time allocated to the study by various

instructors, allowing either one case or two cases to be given out. Because of these factors, as well as the difficulties involved in scheduling additional instructors to participate in the study, the final number of observations for the manager scenario and the grievant scenario were quite uneven with regards to the SEQ-ORDER and the QUES-ORDER variables.

The uneven pattern of observations was more apparent in the data set generated from the grievant scenario, as the SEQ-ORDER the QUES-ORDER variables are not fully crossed. This is due to the fact that the grievant cases containing the second form of the questionnaire (with general items first followed by specific items) could not be given (as the second case in the sequence) after the subjects had responded to a case of the manager. Since the time allocated for the experiment by the instructor determined the number of cases that could be given to the students, and since it was not known in advance which instructors would participate in the experiment and how much time would be allocated by each one of them to the experiment, the unevenness in the data was not surprising. This does not, however, have any substantive implications for interpreting the results from the pilot study. Indeed, the number of observations produced from responses to both cases was quite large. The data, therefore, was amenable to higher level statistical analysis leading to meaningful inferences with regards to how well the manipulations had worked as well

as about the psychometric properties of the scales.

The number of observations greatly exceeded the number of subjects that participated in the experiment. Overall, 161 students were given both cases each, while 136 were given only the case of the manager and 103 were given only the case of the grievant. This produced 297 observations for the manager scenario and 264 observations for the grievant scenario for a total of 561 responses. Due to an overlap of subjects between a managerial applications class and a business strategy class, responses of 7 subjects to the manager scenario and 8 subjects to the grievant scenario had to be eliminated as these subjects had previously responded to a similar case. Furthermore, one additional subject responding to the grievant scenario had not completed the opscan form and hence was dropped from the analysis. A total of 290 observations for the manager scenario and 255 observations for the grievant scenario were available for final tabulations and analysis.

However, out of a total of 290 subjects who had read the case of the manager, only 272 gave usable responses to item 51, which asked them to select one of the 7 decisions at the end of the case. Similarly, out of 255 subjects who had been given the grievant scenario, only 242 gave usable responses to item 51. Therefore the analysis of the decision variable in both the grievant and the manager scenarios uses fewer subjects.

### General Analysis for the Pilot Study

Since the observations were collected using the opscan forms, it was possible to directly code the data and read it into SAS files. This process has the advantage of minimizing human errors associated with data entry. However, the data file has to be looked at carefully to determine whether the respondents simply forgot to code in information which is retrievable from the original opscan forms or the original documentation. For example, two missing values pertaining to gender of the subjects in the manager data set were filled in by referring to the original opscan forms which contained the names of persons. Five missing values in the grievant data and three missing values from the manager data set, all pertaining to the QUES-ORDER variable, were also filled in by going back to original documentation. Furthermore, if for any item a person had marked more than one response, the response was deleted and treated as a missing value.

One of the major reasons for doing the pilot study was to explore the psychometric properties of the two 50-item questionnaires and construct the appropriate dependent variables that would adequately capture how favorably the research subjects evaluated the target person in the work context. Factor-analytic techniques were used to achieve that purpose. The widespread use of this approach has been noted by Kim and Mueller (1978a), who wrote that, "It is not misleading

to argue that the main motivation behind the use of factor analysis is not in ascertaining the factor structures among a set of variables, but in achieving data reduction and obtaining factor scales which can be used as variables in a different study" (p. 51). It was certainly one of the major motivations behind the pilot study to ascertain whether "good" dependent variables measuring research subjects' perceptions of the target person in the work context could be created from the 50-item scales. It was important that the scales constructed to measure the evaluations of the research subjects be reliable.

The analysis performed represents what Comrey (1988) in an important didactic piece described as "the application of tried-and-true exploratory factor-analytic methods to the task of scale development" (p. 760). (See also Comrey, 1973; Devellis, 1991; Kim & Mueller, 1978a, 1978b; and Pedhazur & Schmelkin, 1991.) Indeed, Comrey's (1988) and Pedhazur & Schmelkin's (1991) suggestions and recommendations concerning the use of exploratory factor analysis leading to the selection of particular items to form factor-based scales were adopted virtually in their entirety.

Having a large sample is important in doing an adequate factor analysis (Pedhazur & Schmelkin, 1991). A sample size of 200 is considered minimally satisfactory for most types of factor-analytic techniques, although factor structures are



likely to considerably improve and stabilize with even larger samples (see Comrey, 1988, p. 759). Since the pilot study contained 290 observations in the manager scenario and 255 observations in the grievant scenario, sample size did not pose a problem.

Both the grievant data set as well as the manager data set contained very few missing values in the first 50 items of the questionnaire. There was no evidence from the frequency analysis that certain questions were systematically being omitted. Therefore, the final results were not expected to be biased due to missing data. Nevertheless, to handle the few missing values, a pairwise deletion approach was used. This involved first generating a correlation matrix from the 50 items in the scale and then using it as input to conduct the factor analysis. (If SAS is given raw data as input for factor analysis it simply ignores all observations with any missing values in generating the correlation matrix.)

For both data sets a principal components analysis was performed with varimax rotation. Several factor solutions were attempted using the scree plot as a general guide. In the scree plot, the number of factors is plotted (on the horizontal axis) against eigenvalues (on the vertical axis). This graph indicates the point where successive factors explain relatively little variance, and therefore is helpful in determining the number of factors to extract. (See

Devellis, 1991 for a good discussion of the potential criteria to be used in determining factor solutions).

The suggestion of Harman (1967; 1976) that the objective of factor analysis is achieved when one has an interpretable factor structure was used as the final criterion for accepting the factor solution. This suggestion has much support in the literature and appears in various forms. For example, Devellis (1991) also has stated, "The meaning or interpretability of the resultant factors provide important clues about how many factors to retain" (p. 99).

A cutoff point of .40 was used for item loadings on the factor. Having this cutoff point implied that in order for an item to be included in the factor-based scale, it would have to share at least 16% of the variance with the factor. However, if more than one item loaded onto the same factor at .40 or more, both were deleted from the analysis (with one exception, explained in a later section, pertaining to a factor scale not used in hypotheses testing). This is consistent with the recommendation of Comrey (1988) that an item not be made to do double duty by being in two scales simultaneously. (See also Comrey, 1973; Pedhazur & Schmelkin, 1991; and Kim & Mueller, 1978a for a discussion on the appropriate loading cutoff point in factor analysis.)

After factor-based scales had been determined, they were then subjected to a reliability analysis, and their internal

consistency was determined by calculating the Cronbach's alpha. In further analysis, the intercorrelation of each item to the set of all other items in the scale was determined. Items having the lowest intercorrelations were dropped from the scales and Cronbach's alphas were then recalculated. If the new alphas improved then the dropped items were permanently deleted from the scale. However, if the alpha improved marginally in this process and the scale had only a few items, then Devellis' (1991) rather subtle point was considered before deciding to drop or retain the item. "*The reliability of an alpha as an estimate of reliability increases with the number of items. This means that an alpha computed for a longer scale will have a narrower confidence interval around it than will an alpha computed for a shorter scale. Across administrations, a longer scale will yield more similar values for alpha than will a shorter one*" (p. 89). Of course, if the alpha went down due to deletion of the item, the item was again added back to the scale.

This analysis was performed on each item in all of the factor-based scales. The resulting scales were used as appropriate dependent variables for analysis with all items weighted equally. The decision to give items equal weights in scales has extensive support in the literature and is consistent with the recommendations of Babbie (1990; 1992), Comrey (1988), Nunnally, (1978), and Pedhazur & Schmelkin

(1991). For example, Babbie (1992) states, "I suggest-and practice tends to support this method-that items be weighted equally unless there are compelling reasons for differential weighting. That is, the burden of proof should be on differential weighting; equal weighting should be the norm" (p. 175). Pedhazur & Schmelkin (1991), supporting this perspective, write in their methodology text, "Although differential weights may be applied to separate items (as in assigning weights from a factor analysis in producing factor scores), for most purposes unit weighting (i.e., merely summing up the separate responses, thus weighting each of the items equally) has been shown to produce satisfactory results" (p. 125).

#### RESULTS FROM THE GRIEVANT SCENARIO

The responses of 255 subjects to the grievant scenario were available for analysis in the pilot study. Out of these, 139 were male students while 116 were females. One hundred and seventy-eight of the students received cases in which the questionnaire contained the 40 specific questions first, followed by the 10 general ones. Seventy-seven of the students received cases in which the questionnaire contained 10 general questions first, followed by the 40 specific questions. For one hundred and eighty-four of the subjects the grievant case was either the first or the only case they responded to.

Seventy-one of the subjects responded to the grievant case after having responded to the case of the manager.

A frequency analysis along with the means and standard deviations of the items indicated that the 9-point scale had worked well and adequately captured the variance in the responses of the subjects. Item variance is a valuable attribute for a scale item to have because it allows discrimination among individuals on the construct being measured (Devellis, 1991). It was further evident from the descriptive statistics that there were very few missing values in the data. Extremely few missing values in the data can allow the researcher the option of ignoring them in the analysis (Babbie, 1992).

One hundred and sixty-eight of the 255 students responding to the grievant scenario, because they had been given ample class time, had followed instructions to give reasons for their decision (solicited in item 51). A review of the written comments indicated that the research subjects were concerned about the decision the supervisor made about the target person as well as the actual performance of the target person. The majority of the comments were critical of the supervisor and suggested that he had over reacted in firing the target person, regardless of whether the target person's workplace behavior was aggressive or non-aggressive.

However, the student subjects appeared to be more

positive towards the non-aggressive target person and used generally more favorable words such as "team player" and "helpful" in their description. Many student subjects used slightly negative terms to describe the behavior of the aggressive target person, terms such as "having an attitude" and "poor interpersonal skills."

Since only about 66% of the students wrote comments, no definitive conclusion could be reached regarding the success of the gender-role manipulation in the scenario, although there were indications that it had worked well. Several different statistical analyses based on responses of all 255 students subjects (given later), however, supported the conclusion that gender-role manipulation had been successful.

Since Joe is clearly a male name and Jane is clearly a female name, and since these names occurred quite frequently in the case along with the appropriate pronouns, there was no need to explicitly check for the gender manipulation. Moreover, in reading the written comments of the 168 students, it was noted that the respondents invariably used the name of the target person in their comments (either Joe or Jane), indicating they were quite aware of the gender of the target person.

Many of the written reasons given by the students for the decision made were quite thoughtful. This was an indication that the case had held the attention of the subjects and that

they had found it interesting to read. It was also clear that the 50-item evaluative scale was quite comprehensive in assessing attitudes toward the target person since the comments of the students did not indicate that some additional, overlooked factors pertaining to the case needed to be tapped.

### Constructing the Dependent Variables

A factor analysis with varimax rotation revealed four factors embedded in the 50-item scale. Using the .40 loading as the cutoff point, four different scales based on the factors were constructed. These scales were subjected to reliability analysis and refined further by deleting the variables which were not contributing to the overall reliability. The four factor-based scales are given below.

FACTOR-1 (named GENERAL because general items had high loadings on it. Note that high scores on this factor indicate more favorable evaluation of the grievant).

G8. How would you rate Joe's effort level.

G5. How would you rate Joe's work habits.

G10. What is your overall evaluation of Joe as a worker.

G7. How would you rate Joe's level of work motivation.

G1. How would you rate Joe's overall performance.

G6. How would you rate Joe's innate abilities and aptitudes.

- G3. How would you rate Joe's potential for future promotion.
- S32. Joe works hard at his job.
- G9. How would you rate the contribution of Joe's personality to work place effectiveness.
- S15. Joe is a conscientious worker.
- S9. Joe contributes significantly to the overall department effectiveness.
- S36. Joe puts a lot of energy in his work.
- S27. Joe does not put enough effort in his work (reverse scaled).

FACTOR-2 (named DEC-FAIR to reflect the high loadings of the items addressing the fairness of the decision. Note that high scores on this factor indicate more favorable evaluation of the target person and disagreement with the decision).

- S21. Stanley's decision regarding Joe lacks basic fairness.
- S29. Stanley made a well thought out decision when he fired Joe (reverse scaled).
- S1. Stanley is a fair manager (reverse scaled).
- S38. Joe's present situation is a result of Stanley's unfairness.
- S25. Stanley's decision to fire Joe is an overreaction.
- S18. Joe has no one to blame but himself for his predicament (reverse scaled).
- S37. Joe's mistake reveals something about Joe that is



unlikely to change in the future (reverse scaled).

S12. Joe is careless at work (reverse scaled).

S5. Joe is likely to make serious mistakes at the workplace in the future (reverse scaled).

FACTOR 3 - (Named TYPWORK as the items loading high on it address the nature of work performed. Note that high scores here mean that the target person's work is considered more complex by the research subjects.).

S34. Joe has a demanding job.

S6. Joe's job is an easy one (reverse scaled).

S26. The type of work Joe performs is relatively routine (reverse scaled).

S30. Joe is a highly skilled worker.

S20. The work performed in Joe's department requires specialized training.

FACTOR 4 - (named BACKUP as the highest loading items address the issue of backing-up the data.).

S17. Joe has the responsibility for saving data at the proper time.

S3. Joe did not back up the data when he was supposed to.

S7. Joe gets along well with his co-workers.

S14. Joe is not required to backup the data as part of his job (reverse scaled).

**Descriptive Statistics and Exploratory Analysis**

Descriptive information is provided on the number of observations at each class level of the independent variables.

**TABLE 1**

**THE GRIEVANT SCENARIO-PILOT STUDY  
LEVELS OF THE INDEPENDENT VARIABLES**

<b>RATERSEX</b>	<b>FREQUENCY</b>	<b>PERCENT</b>	<b>CUM-FREQUENCY</b>
FEMALES	116	45.5	116
MALES	139	54.5	255
<b>TARGETSEX</b>	<b>FREQUENCY</b>	<b>PERCENT</b>	<b>CUM-FREQUENCY</b>
FEMALES	130	51.0	130
MALES	125	49.0	255
<b>GEN-ROLE</b>	<b>FREQUENCY</b>	<b>PERCENT</b>	<b>CUM-FREQUENCY</b>
NON-AGGRESSIVE	131	51.4	131
AGGRESSIVE	124	48.6	255
<b>QUES-ORDER</b>	<b>FREQUENCY</b>	<b>PERCENT</b>	<b>CUM-FREQUENCY</b>
FIRST	178	69.8	178
SECOND	77	30.2	255
<b>SEQ-ORDER</b>	<b>FREQUENCY</b>	<b>PERCENT</b>	<b>CUM-FREQUENCY</b>
FIRST	184	72.2	184
SECOND	71	27.8	255

The means and standard deviations of the four factor-based scales as well as the OUTCOME variable are given next. This is followed by a correlation matrix containing the factors, the OUTCOME variable, and the independent variables.

## TABLE 2

### THE GRIEVANT SCENARIO-PILOT STUDY

#### THE MEANS AND STANDARD DEVIATIONS OF FACTORS AND OUTCOME

VARIABLE	N	MEAN	STD-DEV	ALPHA
GENERAL	255	5.6434	1.2089	.94
DEC-FAIR	255	5.8261	1.4605	.88
TYPWORK	255	5.1929	1.1714	.68
BACKUP	255	8.0157	1.0794	.60
OUTCOME	242	3.5744	1.4478	N.A

**TABLE 3****THE GRIEVANT SCENARIO-PILOT STUDY****CORRELATIONS OF FACTORS, OUTCOME, AND INDEPENDENT VARIABLES**

(THE LEVEL OF SIGNIFICANCE IS GIVEN UNDER THE CORRELATION)

	<b>OUTCOME</b>	<b>GENERAL</b>	<b>BACKUP</b>	<b>TYPWORK</b>	<b>DEC-FAIR</b>
<b>OUTCOME</b>	1.00 0.0	-.52801 .0001	.05584 .3871	-.24799 .0001	-.58259 .0001
<b>GENERAL</b>	-.52801 .0001	1.00 0.0	.08432 .1795	.41320 .0001	.69859 .0001
<b>BACKUP</b>	.05584 .3871	.08432 .1795	1.00 0.0	.00741 .9063	-.01554 .8050
<b>TYPWORK</b>	-.24799 .0001	.41320 .0001	.00741 .9063	1.00 0.0	.29866 .0001
<b>DEC-FAIR</b>	-.58259 .0001	.69859 .0001	-.01554 .8050	.29866 .0001	1.00 0.0
<b>RATERSEX</b>	-.07937 .2186	-.01842 .7697	-.06894 .2727	-.02842 .6515	-.10416 .0970
<b>TARGETSEX</b>	-.05594 .3863	.02621 .6770	-.05433 .3876	.04618 .4628	.07331 .2434
<b>GEN-ROLE</b>	.16721 .0092	-.11517 .0663	.02953 .6388	-.02634 .6755	-.17942 .0040
<b>SEQ-ORDER</b>	-.04202 .5154	.09042 .1499	.14526 .0203	-.09204 .1427	.02341 .7099
<b>QUES-ORDER</b>	.07521 .2438	-.11386 .0695	.07301 .2454	-.00918 .8840	-.04266 .4976

A review of the correlation matrix suggests that SEQ-ORDER may be influencing perceptions regarding BACKUP. QUES-ORDER appears to be somewhat correlated to GENERAL, suggesting that if there is a question order effect, general items are being affected the most. This is consistent with the findings of Tourangeau & Rasinski (1988) on question order effects. It is also in line with the assertion of Judd, Smith & Kidder (1991), who suggest that such question order effects generally tend to be weak and certainly weaker than the effects of other factors such as question wording. This does appear to be the case here because almost none of the correlations (with the exception of BACKUP and SEQ-ORDER) involving SEQ-ORDER or QUES-ORDER are significant at the .05 level.

GEN-ROLE is correlated to OUTCOME, DEC-FAIR, and GENERAL, providing statistical evidence, in addition to the qualitative evidence from the written comments of the students, that the manipulation of gender-role was successful. GENERAL and DEC-FAIR are highly correlated, sharing approximately 50% of the variance. This raises some question about the discriminant validity of the two factors. This problem was addressed by comparing the three-factor solution in which GENERAL and DEC-FAIR merge and form one factor with the four-factor solution where GENERAL and DEC-FAIR appear as separate factors.

The reliability of the factor containing both GENERAL and

DEC-FAIR is .94. However, in the four-factor solution where DEC-FAIR appears as an independent factor, the reliability of GENERAL still remains .94. DEC-FAIR, the new factor in the four-factor solution, reaches a reliability of .88. Based on this comparison, it is felt that DEC-FAIR, although highly correlated to GENERAL, is a distinct factor. Furthermore, it is apparent that both QUES-ORDER and SEQ-ORDER have no impact on DEC-FAIR but are weakly correlated to GENERAL. This implies that DEC-FAIR may constitute a less biased and a more appropriate dependent variable to be analyzed in the context of evaluating the grievant.

Another issue that is of some concern from the correlation matrix is that OUTCOME is highly correlated to both GENERAL and DEC-FAIR, the two variables measuring evaluation of the target person. Since the magnitude of the relationship between OUTCOME (which is based on a single item) and GENERAL as well as DEC-FAIR is similar to what one would expect of an item actually belonging to the evaluation scale, the use of OUTCOME as a separate variable is questionable. This is particularly true since the reliability of a single-item variable is inherently uncertain. From a psychometric perspective, single-item scales can pose serious problems in the interpretation of results (Pedhazur and Schmelkin, 1991).

### Statistical Analyses for Manipulation Checks

Using the SAS General Linear Models (GLM) procedure, models for GENERAL, DEC-FAIR, TYPWORK, BACKUP, and OUTCOME were tested (with SEQ-ORDER, QUES-ORDER, RATERSEX, TARGETSEX, and GEN-ROLE as independent variables). The GLM procedure is highly versatile, allows for many different types of exploratory analyses, and can handle both classification and continuous variables. It was used in the pilot study to conduct manipulation checks for GEN-ROLE as well as to estimate the influences of the QUES-ORDER and SEQ-ORDER variables.

As mentioned in an earlier section, because of the unevenness with which SEQ-ORDER and QUES-ORDER could be manipulated in various classes (they are not fully crossed), the models estimated for the factor-based scales do not have interaction terms containing both SEQ-ORDER and QUES-ORDER, and hence are incomplete. This, however, does not influence the substantive results in any meaningful way.

Because of unequal cell sizes, TYPE III sums of squares were obtained for the independent variables. The use of TYPE III sums of squares in unbalanced designs has extensive support in the current statistical literature and is consistent with the recommendations of Howell (1992). This method has gained particular credence in recent years in handling the problems associated with unequal cell sizes in

factorial experiments. TYPE III sums of squares allow a partial F test, which tests for the contribution of a regressor (an independent variable) controlling for all other independent variables. In other words, the variance explained by the independent variable entering the model as the last variable is tested for significance.

In this approach all cell means are weighted equally, and the number of observations in the cell from which the mean is obtained is ignored as an input (Pedhazur, 1982). Both Pedhazur (1982, p. 323) and Howell (1992, p. 544) suggest that this method is particularly appropriate when cell sizes are independent of treatment conditions (in other words when unequal n's are not the result of subject mortality). Many of the statistical packages routinely calculate estimates of treatment effects based on equally weighted cell means when the cell sizes differ (see, for example, SAS, version 6).

Although there is still some disagreement among statisticians as to how unequal cell sizes should best be handled in factorial designs, Howell (1992) summarizes the issues and the debate and comes to the following judgement: "...the conclusion to be drawn from the literature at present is that for most common situations Method 1 is appropriate, since we usually wish to test equally weighted means...Method 1 sums of squares are the values labeled as TYPE III SS in SAS" (p. 545). The pilot study, therefore, based on the



statistical and methodological literature in the area, tested all the models in the exploratory analysis using TYPE III sums of squares.

The results confirmed the findings from the correlation matrix. First, GEN-ROLE entered the model explaining GENERAL at a significant level ( $p < .036$ ), indicating that this manipulation had been successful in that subjects were clearly able to distinguish between grievants who were aggressive and assertive and those who were not. For DEC-FAIR, the effect size was greater, with GEN-ROLE entering the model at a significant level ( $p < .013$ ) providing additional evidence for the success of the manipulation. Finally, for the OUTCOME variable measuring the extent to which the actual decision made by the research subjects (item 51) was favorable to the grievant, GEN-ROLE again entered the model at a statistically significant level ( $p < .003$ ).

Second, QUES-ORDER ( $p < .134$ ,  $p < .62$ ) and SEQ-ORDER ( $p < .53$ ,  $p < .87$ ) did not enter the two models (for GENERAL and DEC-FAIR respectively) at statistically significant levels. For BACKUP, QUES-ORDER ( $p < .65$ ) and SEQ-ORDER ( $p < .07$ ) did not enter the model significantly, although the effect size for SEQ-ORDER was greater. The implication here is that SEQ-ORDER may potentially influence certain types of perceptions. For TYPWORK, neither QUES-ORDER ( $p < .34$ ) nor SEQ-ORDER ( $p < .18$ ) entered the model significantly. Finally, for the OUTCOME

variable, the effect of neither QUES-ORDER ( $p < .33$ ) nor SEQ-ORDER ( $p < .96$ ) was significant.

### Conclusions

The pilot study with the students involving the case of the grievant leads to the following conclusions.

1. Two major factor-based scales pertaining to the evaluation of the target person emerge from the factor analysis. They share approximately 50% of the variance.

2. Both factors have high reliabilities (.94 for GENERAL and .88 for DEC-FAIR).

3. There is both qualitative evidence (from the written comments of the 168 students) as well as statistical evidence (from the correlation matrix and GLM models) that the gender-role manipulation in the scenario was successful. GEN-ROLE enters the model for DEC-FAIR at a higher level of significance than the model for GENERAL.

4. Evidence suggests that DEC-FAIR is less influenced by QUES-ORDER and SEQ-ORDER than GENERAL. OUTCOME is also not influenced by QUES-ORDER or SEQ-ORDER.

5. DEC-FAIR has a slightly higher correlation to OUTCOME than GENERAL.

6. GENERAL shares 16% of the variance with another factor, TYPWORK. DEC-FAIR shares 9% of the variance with TYPWORK.

7. OUTCOME is highly correlated to GENERAL and DEC-FAIR. Since the OUTCOME variable is based on a single item, it brings into question whether OUTCOME should be analyzed separately from DEC-FAIR or GENERAL, both of which have high reliabilities.

One of the purposes of the pilot study was to construct a dependent variable that would adequately capture the variance in the evaluations of research subjects. Based on the results, both GENERAL and DEC-FAIR appear to be suitable candidates for that role as both are highly reliable. The following arguments explore the qualities of both factors.

Any scale measuring the research subjects' evaluations of the target person should correlate highly to the final decision made about the target person, thereby demonstrating convergent validity. At the same time, such a measure should not correlate highly with measures of only weakly related concepts thus demonstrating discriminant validity (see Babbie, 1992; Cook and Campbell, 1979; Kerlinger, 1986; and Pedhazur & Schmelkin, 1991 for good discussions on convergent and discriminant validity).

DEC-FAIR satisfied both criteria better than did GENERAL in the pilot study. In addition DEC-FAIR was not influenced by the methodological manipulations involving QUES-ORDER and SEQ-ORDER, due probably to the fact that it does not contain any general items (see Tourangeau & Rasinski, 1988, who argue that

specific items are less susceptible to order effects). Furthermore, considering that more than half the items in the DEC-FAIR scale in the pilot study are reverse scored, it still achieved a respectable reliability of .88.

On the other hand, GENERAL is a broader measure of the evaluation of the target person because it taps both general and specific items and achieves a reliability of .94, which is somewhat higher than that of DEC-FAIR. While GENERAL appears to be more influenced by QUES-ORDER and SEQ-ORDER than DEC-FAIR, that influence is rather weak and not statistically significant in any of the models. However, according to the literature (Tourangeau & Rasinski, 1988), general items are more susceptible to question order effects. The possibility that a factor-based scale containing too many general items may be biased by extraneous influences must be considered.

Based on this analysis, no definitive conclusion could be reached as to which variable, GENERAL or DEC-FAIR, should be employed as the dependent variable in the actual study. Since, there was a potential for the factor-based scales in the actual study to differ somewhat from the factor-based scales in the pilot study, that decision had to wait until the data from the actual study had been factor analyzed.

### RESULTS FROM THE MANAGER SCENARIO

Two hundred and ninety subjects gave usable responses to the manager scenario in the pilot study. Out of these, 160 were male students and 130 were females. One hundred and eighty-four of the students received cases in which the questionnaire contained the 40 specific questions first, followed by the 10 general ones. One hundred and six of the students received cases in which the questionnaire contained 10 general questions first, followed by the 40 specific questions. For two hundred and twenty-eight of the subjects the manager case was either the first or the only case they responded to. Sixty-two of the subjects responded to the manager case after having responded to the case of the grievant.

Again, a frequency analysis indicated that the 9-point scale had worked well and had adequately captured the variance in the responses of the subjects. One hundred and sixty-four of the 290 students responding to the grievant scenario gave reasons for their decision (solicited in item 51). A review of these written comments revealed that gender-role behavior manipulation had been very successful. As in the grievant scenario, the respondents invariably used the name of the target person in their written comments (either Ted or Kathy), indicating they were aware of the gender of the target person.

Furthermore, the nature of the written comments reflected

clearly that the respondents were able to distinguish between the two types of managerial behaviors operationalized in the vignettes. The comments with regards to the democratic/participative managers (regardless of their gender) were generally much more positive than the comments pertaining to the autocratic/directive managers. This tendency was nearly universal across respondents.

The fact that the gender-role behavior manipulation worked successfully is strongly borne out by the statistical analyses. Again, the 50-item scale proved to be quite adequate in that the students' comments did not indicate that some additional overlooked factors pertaining to the evaluation of the target person needed to be tapped.

### Constructing the Dependent Variables

A factor analysis with varimax rotation revealed three distinct factors embedded in the 50-item scale. All three factor-based scales had high reliabilities. One general factor, however, appeared to dominate, as a 28 item scale with a reliability of .984 measuring evaluation of the target person emerged. The other two factors tapped the perceptions of the research subjects pertaining to the effort (EFFORT) the target person was making and the role chance or merit (LUCK/MERIT) had played in the advancement of the target person.

The loading of .40 was used as the cutoff point to select items for the three scales, with one exception. One item (S29) that loaded highly (.68) on the LUCK/MERIT factor also loaded at .42 on the EFFORT factor. Since including this item raised the reliability of the scale from .77 to .84, the item was retained in the factor-based scale measuring LUCK/MERIT but not EFFORT. These scales were subjected to reliability analysis and refined further by deleting the variables which were not contributing to the overall reliability. The items contained in the three factor-based scales are given below.

FACTOR 1 (GENERAL EVALUATION FACTOR)-High scores indicate more favorable evaluations of the target person.

S19. Ted is the kind of manager one can go speak openly with.

S18. Ted values participative decision making.

S13. Ted encourages his subordinates to express their views.

S32. Ted has good interpersonal skills.

S10. Ted does not treat his staff with respect (reverse scaled).

G2. How would you rate Ted's interpersonal skills.

G4. How would you assess Ted's value as a team player.

S26. Ted is a fair and just manager.

S17. Ted knows how to get along with people.

S37. Ted knows how to win the support of his subordinates.

S11. Ted considers the opinions of his subordinates before

making decisions.

S21. Ted can be described as a democratic leader.

S1. Ted acts fairly towards his staff.

S5. Ted cares for the people who work for him.

S39. Ted's personality is a great asset for him as a manager.

S27. Ted knows how to effectively manage people.

G8. How would you rate the contribution of Ted's personality to work place effectiveness.

S14. Ted's management style will hinder his future success (reverse scaled).

S34. Ted has an effective management style.

S28. Working for Ted is probably quite stressful for his subordinates (reverse scaled).

G10. How would you rate Ted as a manager.

G9. How would you rate Ted's overall leadership abilities.

S6. Ted is not handling his present job too well (reverse scaled).

S20. Ted does not have the personality to be a successful general manager (reverse scaled).

G3. How would you rate Ted's potential for future promotion.

S38. Ted's subordinates probably perceive Ted to be a competent leader.

S15. Ted has been an effective manager at his present job.

S24. Ted does not have the emotional maturity to be a general manager (reverse scaled).



FACTOR-2 (EFFORT)-High scores indicate a perception by the research subjects of a higher level of effort on part of the target person.

G7. How would you rate Ted's effort level.

S22. Ted puts a lot of energy in his work.

S8. Ted has a difficult task ahead of him.

S25. Ted's job as the acting general manager is demanding.

S16. Ted works hard at his job.

S9. Ted is competent in what he does.

G6. How would you rate Ted's level of work motivation.

S31. Ted successfully accomplishes his goals.

FACTOR 3 (LUCK/MERIT)-High scores indicate that there is a perception of success due to factors internal to the target person. Low scores indicate perceived success due to chance or luck.

S40. Ted's past success was caused by factors outside of Ted (reverse scaled).

S33. Ted's rise to the position of acting general manager has more to do with chance than with merit (reverse scaled).

S35. The cause of Ted's career advancement lies in Ted himself.

S29. Ted's progress through the managerial ranks is no fluke.

Descriptive Statistics and Exploratory analysis

Descriptive information is provided on the number of observations at each class level of the independent variables.

**TABLE 4**

**THE MANAGER SCENARIO-PILOT STUDY  
LEVELS OF THE INDEPENDENT VARIABLES**

<b>RATERSEX</b>	<b>FREQUENCY</b>	<b>PERCENT</b>	<b>CUM-FREQUENCY</b>
FEMALES	130	44.8	130
MALES	160	55.2	290
<b>TARGETSEX</b>	<b>FREQUENCY</b>	<b>PERCENT</b>	<b>CUM-FREQUENCY</b>
FEMALES	148	51.0	148
MALES	142	49.0	290
<b>GEN-ROLE</b>	<b>FREQUENCY</b>	<b>PERCENT</b>	<b>CUM-FREQUENCY</b>
NON-AGGRESSIVE	142	49.0	142
AGGRESSIVE	148	51.0	290
<b>QUES-ORDER</b>	<b>FREQUENCY</b>	<b>PERCENT</b>	<b>CUM-FREQUENCY</b>
FIRST	184	63.4	184
SECOND	106	36.6	290
<b>SEQ-ORDER</b>	<b>FREQUENCY</b>	<b>PERCENT</b>	<b>CUM-FREQUENCY</b>
FIRST	228	78.6	228
SECOND	62	21.4	290

The means and standard deviations of the three factor-based scales as well as the OUTCOME variable are given below. They are followed by a correlation matrix containing the factors, the OUTCOME variable, and the independent variables.

**TABLE 5**

**THE MANAGER SCENARIO-PILOT STUDY**

**THE MEANS AND STANDARD DEVIATIONS OF FACTORS AND OUTCOME**

VARIABLE	N	MEAN	STD-DEV	ALPHA
GENERAL	290	4.5960	2.1247	.98
EFFORT	290	6.8721	1.0153	.87
LUCK/MERIT	290	6.0120	1.5325	.84
OUTCOME	272	2.9154	1.3945	N.A

**TABLE 6****THE MANAGER SCENARIO-PILOT STUDY****CORRELATIONS OF FACTORS, OUTCOME AND INDEPENDENT VARIABLES**

(THE LEVEL OF SIGNIFICANCE IS GIVEN UNDER THE CORRELATION)

	<b>OUTCOME</b>	<b>GENERAL</b>	<b>LUCK/MERIT</b>	<b>EFFORT</b>
<b>OUTCOME</b>	1.00 0.0	-.57513 .0001	-.44319 .0001	-.37263 .0001
<b>GENERAL</b>	-.57513 .0001	1.00 0.0	.40243 .0001	.36005 .0001
<b>LUCK</b>	-.44319 .0001	.40243 .0001	1.00 0.0	.52769 .0001
<b>EFFORT</b>	-.37263 .0001	.36005 .0001	.52769 .0001	1.00 0.0
<b>RATERSEX</b>	-.01717 .7780	-.02453 .6775	-.06805 .2480	-.02686 .6488
<b>TARGETSEX</b>	.00658 .9139	.00334 .9548	-.03854 .5133	-.03430 .5608
<b>GEN-ROLE</b>	.40366 .0001	-.82266 .0001	-.19254 .0010	-.19952 .0006
<b>QUES-ORDER</b>	.07802 .1996	-.06153 .2963	-.08438 .1518	-.04685 .4267
<b>SEQ-ORDER</b>	-.07056 .2461	.10158 .0842	.09668 .1004	.06912 .2406

A review of the correlation matrix indicates that SEQ-ORDER and QUES-ORDER are not significantly correlated to any of the factors or OUTCOME. There is some suggestion of a weak correlation between SEQ-ORDER and GENERAL.

GEN-ROLE, on the other hand, is highly correlated to both OUTCOME and GENERAL, providing strong statistical evidence in addition to the qualitative evidence from the written comments of students that the manipulation of gender-role was successful. GENERAL and LUCK/MERIT are correlated, sharing approximately 16% of the variance. GENERAL and EFFORT are also correlated, sharing about 13% of the variance.

Notably, the OUTCOME variable based on one item is highly correlated to GENERAL. This correlation is of the magnitude that one would expect of an item actually belonging to the evaluation scale. This calls into question the utility of analyzing the OUTCOME variable separately from GENERAL in the manager scenario as well, particularly since GENERAL has a high reliability and the reliability of OUTCOME is uncertain.

### **Statistical Analyses for Manipulation Checks**

Using the SAS GLM procedure, models for GENERAL, LUCK/MERIT, EFFORT, and OUTCOME were tested (with SEQ-ORDER, QUES-ORDER, RATERSEX, TARGETSEX, and GEN-ROLE as independent variables). Because of unequal cell sizes, TYPE III sums of squares were used for analysis. The logic behind the use of

TYPE III sums of squares was given in the earlier section and will not be repeated here. The results of this analysis confirmed the findings from the correlation matrix.

First, the GEN-ROLE entered the model for GENERAL at a significant level ( $p < .0001$ ). For the OUTCOME variable (item 51), measuring the extent to which the actual decision made by the research subjects was favorable to the manager in the case, GEN-ROLE again entered the model at a high level of statistical significance ( $p < .0001$ ). Subjects were clearly able to distinguish between the two different management styles (Autocratic v. Democratic) and behaviors.

Second, although QUES-ORDER ( $p < .95$ ) did not enter the model for GENERAL at a significant level, SEQ-ORDER was statistically significant ( $p < .045$ ). This confirms the weak association noted earlier in the correlation matrix between SEQ-ORDER and GENERAL. For EFFORT, QUES-ORDER ( $p < .81$ ) and SEQ-ORDER ( $p < .15$ ) did not enter the model significantly. For LUCK/MERIT, neither QUES-ORDER ( $p < .46$ ) or SEQ-ORDER ( $p < .083$ ) entered the model significantly, although the effect size for SEQ-ORDER was larger. Finally, for the OUTCOME variable, the effect of neither QUES-ORDER ( $p < .28$ ) nor SEQ-ORDER ( $p < .35$ ) was significant.

### Conclusions

The pilot study with the students involving the case of the manager leads to the following conclusions.

1. Three highly reliable factor-based scales emerge from the data. The GENERAL scale shares 13% of the variance with EFFORT and about 16% of the variance with LUCK/MERIT. LUCK/MERIT and EFFORT share approximately 27% of the variance.

2. There is both qualitative as well as statistical evidence indicating that the GEN-ROLE manipulation was successful.

3. SEQ-ORDER is weakly correlated to GENERAL and LUCK but enters the model at a significant level only for GENERAL ( $p < .05$ ).

4. As in the case of the grievant, QUES-ORDER does not enter any of the models at a statistically significant level.

5. The OUTCOME variable is highly correlated to GENERAL, calling into question whether it should be treated as a separate variable. This is of particular concern since the OUTCOME variable is based on one item whose reliability is uncertain.

Based on this analysis, it can be said that the GENERAL factor, with a reliability of .98, is a highly suitable dependent variable for testing proposed hypotheses pertaining to evaluation of the target person in the manager scenario. However, SEQ-ORDER may have an influence on GENERAL if the subjects were given two cases to read. This issue needed to be addressed.

### REFINING THE DESIGN BASED ON THE RESULTS OF THE PILOT

The pilot study allowed for psychometric and exploratory analysis that was helpful for refining the design of the actual study. It was found that the GEN-ROLE manipulation was successful in both the cases. The written comments of the students indicated that the cases held their interest and attention. It was also clear that the 9-point scale worked as intended and that the scale-related instructions were easy to follow.

However, it was noted that several people (18 out of 290 in the manager scenario and 13 out of 255 in the grievant scenario) in both of the cases either misunderstood the directions for item 51 (asking them to choose one of the 7 decision options) or did not respond to it. Because of this, it was decided to emphasize these scaling instructions more and include an additional page with the 9-point scales and the 7-point decision scale drawn on it to guide the subjects in filling in the opscan forms. This additional page also repeated instructions for responding to item 51. It was hoped that this refinement in scaling instructions would lead to an increase in the accuracy of responses to item 51.

The pilot study further demonstrated that reliable dependent measures of evaluation could be constructed out of the 50-item scales in both of the questionnaires. It was found that the influence of QUES-ORDER was weak at best and that it



did not enter any of the models in the grievant or the manager cases at a statistically significant level. This variable, therefore, was dropped from further consideration.

SEQ-ORDER, however, appeared to have some effect on GENERAL in the manager case and entered the model at a statistically significant level ( $p < .05$ ). Since it was necessary in the actual study to give the subjects both the grievant and the manager scenarios (in order to effectively double the sample size in both data sets), it was decided to include SEQ-ORDER in the design of the study.

In this design, SEQ-ORDER, along with RATERSEX, TARGETSEX, and GEN-ROLE, is treated as an independent variable. It was reasoned that SEQ-ORDER could always be either dropped from the analysis or an adjustment could be made for it in calculating means. The pilot study, therefore, enabled a reconceptualization of the project as consisting of two experiments with the same subjects where the sequence in which the case is given (either first or second) is included in the design as a manipulation to potentially explain the responses of the subjects.

The major advantage of including the SEQ-ORDER variable in the model was that it allowed two cases (both the manager and the grievant case) to be given to the same subjects. From a practical standpoint, this doubling of the number of observations in each of the two samples was important because

it allowed for an increase in statistical power.

In addition to the above refinements, it was also decided to use the randomized block design in the actual study with the practitioners (with the sex of the subjects as the block). This was important as the sample of managers for the study was expected to be smaller than the sample of students. In the larger student samples, even though males and females differed in their relative proportion (45 to 55), simple random assignment was sufficient to ensure that the numbers of males and females in each of the cells would be adequate.

However, for a smaller sample with a more extreme difference in the relative proportions of men and women, simple random assignment could lead to highly unequal cell sizes, making statistical analysis complex and results more difficult to interpret. Randomized block design increases the probability that cell sizes will be in rough proportion to the proportion of males and females in the sample. This can serve to reduce the error term and make the statistical analysis more efficient.

#### **CHOICE OF THE SITE, PLANNED ANALYSIS**

#### **AND DATA COLLECTION PROCEDURES**

To conduct an on-site study requires access to a cooperative organization, a requirement which may at times be problematic (see Cook & Campbell, 1979). Often it is difficult

to convince top management that they have something to gain by allowing the study to take place. The on-site study, however, can be valuable if the research is supported by top management and if the subjects participate in it with a spirit of enthusiasm and cooperation. Fortunately, in this study, all these conditions were met to a large extent.

For the purposes of conducting this study, the director responsible for training and development in the Virginia Cooperative Extension (VCE) was contacted with a request for using unit directors in the sample. After some discussion, it was formally proposed to the top management of VCE that this on-site experimental study would give valuable insight into the type of management styles considered acceptable within the extension, the nature of management evaluation and decision making, as well as the potential for gender bias or sex discrimination in the organization due to the judgment of managers.

Since the director of extension was already considering training and development of unit directors with respect to evaluation judgements and workplace diversity issues, the proposal proved to be timely. After the director reviewed the cases and the questionnaires constructed for the study, a process was initiated to convince other top managers in the VCE that this project was worth pursuing.

Eventually an agreement was reached that the research

results of the study would be shared with the VCE and appropriate feedback would be given to the unit directors during their regional meetings. Such results could potentially be used to institute training programs to sensitize unit directors to various management style issues as well as issues pertaining to gender bias in evaluation and decision making. The study, therefore, took place under the auspices of the training and development of unit directors and served the mutual interests of the VCE and the researcher.

No changes were suggested in the instruments by the top management, except for the addition of a demographic questionnaire. The VCE wanted certain information collected concerning the unit directors and the districts that they belonged to. The demographic portion of the questionnaire represented an attempt to gather that information for the VCE.

### The Theoretical and Practical

#### Implications of the Sample

Combining the research of theoretical issues with practitioner concerns by conducting field studies has been suggested often by management scholars (Thomas and Tymon, 1983). As noted by Kerlinger (1986), "Field experiments are well suited both to testing theory and to obtaining answers to practical questions. Whereas laboratory experiments are suited mainly to testing aspects of theories, field experiments are

suites both to testing hypotheses derived from theories and to finding answers to practical problems" (p. 379). This study did allow for the best of both worlds in selection of the sample. The choice of this particular sample was theoretically important and methodologically relevant.

First, the unit directors all belonged to the same organization and in that respect could be considered a homogeneous group of managers. Second, since the research took place on-site, the potential problem of non-response was no longer an issue. Third, typically among a sample of managers, the overwhelming majority tend to be male. This leads to difficulties in interpreting results pertaining to possible differences between male and female managers (research subjects) in how they evaluate or make decisions. Among the unit directors in the VCE, approximately 38% were females. Fourth, the support of top management for this project assured that ample time would be allocated for doing the study and that there would be enthusiastic participation and involvement by virtually all of the unit directors in four districts of the state of Virginia. The satisfying of these important criteria created ideal conditions for conducting a field experiment with this group of managers.

### Planned Statistical Analysis to Test the Hypotheses

The data collected using the experimental methodology is

amenable to many different types of analysis (such as ANOVA or REGRESSION). However, the hypotheses advanced involve specific, planned non-orthogonal comparisons. For testing planned comparisons, the F test of ANOVA, which addresses the more general question of whether the category means are equal, is unnecessary and in fact irrelevant (Howell, 1992; Pedhazur & Schmelkin, 1991). The following procedure involving multiple comparisons between the means was planned in advance as the most appropriate for testing the proposed hypotheses. The logic underlying the statistical analysis is outlined below.

Since the interest of the study was on examining the differences between the cells only as they were dictated by theoretical considerations, the most precise method of analysis involves a priori comparisons. A priori contrasts are chosen before the data is collected and therefore must be distinguished from post hoc comparisons. Post hoc comparisons are planned by the researcher after the data has been collected and examined carefully. Such an approach, because it capitalizes on chance to find significant results, can lead to a dramatic increase in the Type 1 error (see Howell, 1992). However, this possibility is minimized when, through the development of a conceptual framework, specific contrasts can be planned in advance. Planned contrasts allow the use of more powerful statistical approaches relative to the rather conservative procedures such as Scheffe's that must be relied

on for post hoc comparisons (see Howell, 1992 or Pedhazur & Schmelkin, 1991 for an excellent discussion of both planned and post hoc multiple comparison methods).

Because the planned contrasts in this study are not orthogonal (see Howell, 1992 for a discussion of orthogonal contrasts), the technique known as Bonferroni t statistics (Miller, 1966) was considered to be the most suitable for analyzing the differences between the cell means. In this approach, the overall alpha level (Type 1 error) is divided by the number of planned non-orthogonal comparisons to derive an effective alpha. In other words, the alpha level is adjusted depending on the total number of t tests to be performed (see Pedhazur & Schmelkin, 1991). The decision was made to maintain overall alpha level at the conventional .05 level of significance for the complete set of hypotheses. Since there were five research hypotheses, this implied that the differences in the appropriate cell means would be tested at the .01 level of significance.

### **The Influence of Cell Size on the Analyses and Interpretation**

Both the experiments (involving the grievant and the manager) were expected to result in 16 cells each, as all four of the independent variables (SEQ-ORDER, RATERSEX, TARGETSEX, GEN-ROLE) are measured at two levels. If the SEQ-ORDER variable turned out not to be significant in the analysis and

was dropped, it would result in 8 cells for each of the experiments, which could then be compared to test the hypotheses postulated earlier.

Since the number of male and female subjects was not equal in the population of the unit directors, it was evident that the cell sizes in both of the designs would differ. This was of some concern because having equal number of subjects in each cell is important to maintaining the orthogonality in a factorial experimental design such as the one proposed (Tabachnick & Fidell, 1989). Equal cell sizes have two important advantages. First, statistical tests such as the t test are more sensitive to differences between the means when the cell sizes are equal. Second, if there are departures from the assumptions underlying statistical tests such as the t, having equal cell sizes helps to minimize the resulting distortions (Pedhazur, 1982).

However, the number of subjects in the cells may be different either by design or due to subject mortality, and it is crucial to understand that difference in interpreting the results. As Pedhazur (1982) has stated clearly, "The use of unequal n's by design does not pose a threat to the internal validity of the experiment, that is, to the valid conclusions about the effect of the treatment. Subject mortality, however, may pose very serious threats to internal validity" (p. 317). In the current study, the greatest threat to the methodology



was posed not by unequal  $n$ 's in the cells per se but only if the non-response of subjects resulted in unequal  $n$ 's in the cells.

From a purely statistical perspective, there are many ways to analyze designs with unequal cell sizes. If, by chance, group sizes differ only slightly, they can be equalized by eliminating subjects from the larger groups at random. This procedure can facilitate data analysis by maintaining orthogonality between the independent variables. This procedure is, however, subject to the criticism that it may unnecessarily lead to loss of data. Furthermore, this method may not be practical if the cell sizes are quite disparate. A second approach called a proportional solution (see Howell, 1992) can be used if the cell frequencies in rows and columns are in some type of proportion to each other. Howell (1992) has emphasized, however, that in this approach, "The sample sizes are treated as part of the treatment effect in that large samples carry more weight in the analysis" (p. 413). In this procedure, then, the analysis takes place on a set of weighted means. A third approach is to simply compare unweighted means. This may be appropriate if the differences in cell sizes represent differences that actually exist in the populations (see Pedhazur, 1982, for that line of reasoning).

In addition to these various methods, the least square means solution has been advocated as a more refined solution

for the problem of unequal cells in the design. In this approach, the unweighted means are adjusted for the purposes of testing the hypotheses pertaining to multiple comparisons between the means. This approach has gained more acceptance among statisticians in recent years (Searle, 1987). Currently many statistical packages, including SAS, calculate least square means. This allows an easy check on how drastically (or how little) the unweighted means change when they are adjusted and allows an inference regarding how much impact the design with unequal cell sizes is having on substantive results.

The concept of least square means is not yet widely found in the social science literature but is well established in agricultural science and statistical journals (for examples, see Searle, Speed, and Milliken, 1980). Unfortunately the term itself is rather misleading. Searle et al. (1980) state, "As it stands the expression *least square means* has no implicit meaning, because the words *least squares* do not function informatively as an adjective to the mean" (p. 216). The SAS/STAT user's guide (version 6, 4th edition) defines least square means as the "expected value of class or subclass means that you would expect for a balanced design involving the class variable with all covariates at their mean value" (p. 948). The use of least square means is recommended by statisticians for comparisons of means in an unbalanced design (Searle, 1987; Searle et al., 1980).

Therefore, based on the most current statistical literature on how to approach designs with unequal cell sizes, it was decided to use the least square means to test the proposed hypotheses in the present study. If the design was seriously unbalanced, the least square means would be quite different from the unweighted means. However, if the unbalanced design was having little or no impact, least square means would be about the same as the unweighted means. Using the least square means allows both a more comprehensive understanding of the impact of the design on the results and a more precise test of hypotheses.

#### Data Collection Procedures

Data was collected in five site visits. Four of the site visits involved field trips to the regional district meetings of the unit directors in the four regions of Virginia. These visits were to Lynchburg (central district), Luray (northern district), Jamestown (eastern district), and Abingdon (southwestern district). There was a high level of enthusiasm and cooperation by the unit directors in all of the regions. About an hour and a half was reserved on the schedule at the regional meetings of the unit directors for this project.

The fifth site visit involved collecting data from 16 higher-level administrators comprising the lead team at a meeting at Virginia Tech. Due to time constraints, the visit

with the lead team was shorter than the visits to the regional meetings of unit directors. Therefore, the subjects were only able to complete one case at the site. Everyone in the lead team was given a second case and requested to mail it in at their convenience.

Each session of the experiment began with an introduction of the researcher by a district director or a higher-level administrator. The researcher was typically introduced as a Ph.D student at Virginia Tech doing research on management evaluation and decision making. The subjects were told by the administrator that they would be engaging in some exercises helpful for making them aware of the issues managers may face in different work situations. The session was typically labeled a management development or training session on the agenda, which set the tone for the experiment. The subjects were made aware at the onset that the results of the study would be shared with them at future meetings of the unit directors.

After the introductions, the researcher typically told the subjects that they would be given two cases to read in which they would evaluate the key person in the case and then make a work-related decision about that person. The cases were then passed out to the subjects, along with an opscan form and an extra sheet containing visual scales and instructions on how to complete them. The version of the case was pre-coded on

the opscan form.

After the cases were passed out, the subjects were told to read the case and then answer the items on the opscan form accompanying them. They were also requested, if they had time at the end, to write some reasons for choosing a particular decision (solicited in item 51) with regards to the target person. The scaling instructions were read out loud, and the subjects were requested to ask questions if there were any ambiguities.

The subjects were reminded that, although it was voluntary, it would be appreciated if they could fill out the demographic data sheet requested by the management of VCE. The subjects were asked to code in their gender and the sequence order of the case on the opscan before starting. This instruction was again repeated at the end of the session before the case was handed back. The cases and the opscan were handed back together under the careful scrutiny of researcher and separated later after verifying that the sequence order as well as the case version had been properly coded. Both the questionnaire and the opscan were given the same number by the researcher when the two were separated for future reference.

Randomized block design was used to assign cases to subjects at four of the regional sites where data was collected from the unit directors. At these sites, experimental subjects were blocked into two groups, males and

females. Male subjects were assigned one of the four versions of the grievant or manager vignettes. This process was the same for female subjects. Once the first case had been completed and handed in, the process was repeated once again with the second case.

At the central district meeting the grievant case was given to the unit directors first. After all the subjects had completed the case and the accompanying questionnaire and had handed it in, they were given the manager case to read and complete. The sequence in which the cases were given was reversed for the northern district meeting. This process was followed until the end of the study. In all, at two of the sites, the grievant scenario was read first and completed, followed by the manager scenario. At the other two sites this sequence was reversed. For the lead team comprised of higher-level administrators the sequence was mixed, with some administrators getting the grievant case first and others getting the manager case first.

The whole data collection process took approximately two months and was spread over March and April of 1993. Based on personal interactions with individuals at the site, it was the impression of the researcher that the subjects enjoyed working on the cases and that the data collected was of good quality. No unusual incidents took place during the data collection and both the management of the extension and the subjects

cooperated fully at all times. Once the experiment was complete, the opscan forms were read by the scanner, and two data files (one for the grievant case and one for the manager case) were constructed. These two data sets were analyzed to test the proposed hypotheses.

## **CHAPTER 4**

### **RESULTS**

#### **INTRODUCTION**

This chapter presents the findings from the tests of the hypotheses with the unit directors. First, important sample characteristics are identified. Second, the factor analysis results and their implications for the reliability of the evaluation measure are discussed for the grievant scenario. Third, the descriptive statistics pertaining to the evaluation factor that emerged from the grievant data set are outlined. Fourth, the results from testing the hypotheses pertaining to the grievant case are presented. Steps two, three, and four are then repeated for the manager scenario, thus concluding the results section.

#### **SAMPLE CHARACTERISTICS AND MEASUREMENT**

At the time that this study was conducted (during March and April of 1993), a total of 134 managers worked for the Virginia Cooperative Extension in the four different geographical regions of the state. Out of these, 116 held the rank of unit director, while 18 held higher administrative ranks. Out of 116 unit directors, 113 were present at the sites where the study took place. Sixteen out of 18 higher ranking managers were also present at one of the sites and



were able to participate in the study. Over 96% of the total population of Virginia Cooperative Extension managers was, therefore, represented in the sample of this study.

Because the subjects had to code in their gender on the opscan form itself, perfect information was available as to the number of male and female managers participating in the field experiment. Out of the 129 subjects, 80 (62%) were males and 49 (38%) were females. Information with regard to other demographic characteristics such as age, race, and education, however, was less complete as filling out the last portion of the questionnaire had been emphasized as voluntary.

The demographic profile, generated from questions requested by the Virginia Cooperative Extension (VCE) to be included at the end of the questionnaire, revealed the following information about the subjects. Eighty-nine percent of the subjects reported they were white while 11% reported they were black. Twenty of the subjects did not indicate their race. Thirty-eight percent of the subjects reported having a bachelor's degree, 56.4% reported having a Master's degree and 6.4% reported having a Ph.D. Seventeen subjects did not indicate their level of education.

Approximately 21% of the subjects ranged in age from 26 to 40. Sixty-six percent of the subjects were between the ages of 41 and 55. Thirteen percent of the subjects were over 55. Twenty-five of the subjects did not indicate their age. Fifty-

one percent of the subjects reported having a tenure of six years or less in their position of unit director. Thirty-nine percent reported having between 7 and 19 years of experience. Ten percent reported having 21 or more years of experience. Thirty-five of the subjects left that portion of the demographic questionnaire blank.

### Subjects' Response Rate

One hundred and twenty-nine unit directors or administrators holding a higher rank in the VCE and working for the state of Virginia participated in this study. Twenty-six of the unit directors were from the central district, 31 of the directors were from the northern district, 36 of the directors were from the eastern district, and 20 directors were from the southwestern district. Sixteen of the subjects held administrative ranks higher than that of unit director and made up the lead team.

All 129 subjects responded to the grievant scenario for a 100% response rate generating 129 observations for that data set. One hundred and twenty-seven out of 129 subjects responded to the case involving the manager, resulting in a response rate of 98.4%. The two non-respondents were male managers belonging to the lead team (higher-level administrators).

### Missing Values in the Data

A review of the original data file revealed that, with the exception of demographic information (which was provided on a voluntary basis), both of the data sets were remarkably complete, with only a few missing values. Some additional missing values were generated by deleting responses to those items where subjects had marked the opscan form twice for the same item. The following descriptive information indicates the overall level of completeness of both the data sets.

G-DATA	TOTAL VALUES POSSIBLE	MISSING VALUES	RATIO-MV/TVP
N=129	129*51=6579	18	.0027
M-DATA	TOTAL VALUES POSSIBLE	MISSING VALUES	RATIO-MV/TVP
N=127	127*51=6477	19	.0029

It is important to note that missing values constituted less than .03% of all the values that were possible given the number of subjects and the 51-item questionnaire. Babbie (1992) suggests that when the number of missing values in a data set is quite small, the missing values can safely be ignored in the data analysis. Although the pairwise deletion approach was used when missing values were encountered, observations containing missing values were ignored for calculating reliabilities of the factor-based scales. Since,

the number of missing values was small, reliability estimates were not expected to be biased by this procedure.

### Constructing Measures Using Factor Analysis

Although the pilot study with the grievant case (255 observations) and the manager case (290 observations) had revealed the factor structures of the two 50-item scales, it was felt that a separate factor analysis was needed for the two data sets generated from the unit directors. It simply could not be assumed that the unit directors would evaluate and judge the target person in the exact same way as the students and respond in an identical fashion to the items. On the other hand, since both the students and the unit directors responded to identical cases and questionnaires, it was expected that there would be broad similarities in the factor structure.

The sample for the unit directors was almost half (in the case of the grievant) or less than half (for the case of the manager) of the student sample. Factor analysis on small data sets has been questioned in the statistical and methodological literature (see Comrey, 1988; Devellis, 1991). As Devellis (1991) points out, however, the smaller the number of factors expected, smaller the sample size can be. Since an extensive pilot study using large samples had already been conducted with identical measures, it was felt that the results from it

(with regards to the number of factors to be extracted) could be used to guide the interpretation of the factor structure found in the data sets generated from the unit directors. The following analysis, therefore, mirrors the analysis conducted in the pilot study. The theoretical and methodological justifications for this approach were discussed in considerable detail in the methodology chapter in the context of analyzing the pilot study data. Consequently, the appropriate citations from the literature will not be repeated here in their entirety.

As in the pilot study, for both data sets, a principal components analysis was performed with varimax rotation on items 1 through 50. Factor solutions were attempted using the scree plot as a general guide. In the scree plot, the number of factors is plotted (on the horizontal axis) against eigenvalues (on the vertical axis). This graph indicates the point where successive factors explain relatively little variance and therefore is helpful in determining the number of factors to extract. (Also see Zwick & Velicer, 1986; and Devellis, 1991, for a good discussion of the potential criteria to be used in determining factor solutions.) As before, the suggestion of Harman (1967; 1976) that the objective of factor analysis is achieved when one has an interpretable factor structure, was used as the final criterion for accepting the factor solution. (See also,

Devellis, 1991, p. 99 for a similar line of argument.)

With one exception (explained later), a cutoff point of .40 was used for item loadings on the factor. Having this cutoff point implied that in order for an item to be included in the factor-based scale, it would have to share at least 16% of the variance with the factor (See also Comrey, 1973; 1988; Pedhazur & Schmelkin, 1991; and Kim & Mueller, 1978a for a discussion on the appropriate loading cutoff point in factor analysis.)

Factor-based scales were determined on this basis and subjected to reliability analysis. Cronbach's alpha for each of the scales was then calculated to check for the reliability of the measures. In the ensuing analysis, where the factor-based scales pertaining to evaluation are used as dependent variables, all items are given equal weight. This approach has considerable support in the methodological and the scaling literature and was discussed in more detail in the methodology chapter. (See for example, Babbie, 1990; 1992; Comrey, 1988; Nunnally, 1978; and Pedhazur & Schmelkin, 1991.) Next, the results pertaining to factor-based scales are given for both the grievant and the manager scenario.

#### **RESULTS FROM THE GRIEVANT SCENARIO**

A factor analysis with varimax rotation revealed four distinct factors embedded in the 50-item scale. The factor

structure in this data set is broadly similar to the data set generated in the pilot study by the students. There were, however, some differences. Using the .40 loading as the cutoff point, four different scales based on the factors were constructed. Because of general similarities, the names of the factors have been kept essentially the same as the names of factors in the pilot study.

FACTOR-1 (This factor was named GENERAL because general items had high loadings on it. Note that high scores on this factor indicate a more favorable evaluation of the grievant).

G10. What is your overall evaluation of Joe as a worker.

G1. How would you rate Joe's overall performance.

G9. How would you rate the contribution of Joe's personality to work place effectiveness.

G7. How would you rate Joe's level of work motivation.

G4. How would you rate Joe's value as a team player in the department.

G3. How would you rate Joe's potential for future promotion.

G2. How would you rate Joe's interpersonal skills.

G5. How would you rate Joe's work habits.

G6. How would you rate Joe's innate abilities and aptitudes.

S2. Joe has the ability to be a good performer.

FACTOR-2 (This factor was named DEC-FAIR/PERFORM to

reflect the high loadings of the items addressing the fairness of the decision and the workplace performance of the target person. Note that high scores on this factor indicate more favorable evaluation of the target person and disagreement with the decision).

S21. Stanley's decision regarding Joe lacks basic fairness.

S16. Joe does not come up to the standards of the department (reverse scale).

S24. Joe has good work habits.

S25. Stanley's decision to fire Joe is an overreaction.

S11. Joe's future performance is likely to be good.

S37. Joe's mistake reveals something about Joe that is unlikely to change in the future (reverse scale).

S12. Joe is careless at work (reverse scale).

S1. Stanley is a fair manager (reverse scale).

S13. Joe successfully accomplishes the tasks assigned to him.

S23. Joe is competent in what he does.

S8. Joe does not have an aptitude for detail work (reverse scale).

S19. Joe is easily distracted from his work (reverse scale).

S5. Joe is likely to make serious mistakes at the workplace in the future (reverse scale).

S28. What happened to Joe could have happened to anybody.

S15. Joe is a conscientious worker.

S29. Stanley made a well thought out decision when he fired



Joe (reverse scale).

S18. Joe has no one to blame but himself for his predicament (reverse scale).

FACTOR 3 - (This factor was named TYPWORK/EFFORT as the items loading high address the nature of the work performed as well as the target person's level of effort. Note that high scores here mean that the target person's work is considered more complex and that the target person is perceived to be exerting a higher level of effort).

S34. Joe has a demanding job.

S36. Joe puts a lot of energy in his work.

S33. Joe is helpful at the work place.

S10. The cases assigned to Joe are complex.

S32. Joe works hard at his job.

S30. Joe is a highly skilled worker.

S9. Joe contributes significantly to the overall department effectiveness.

S20. The work performed in Joe's department requires specialized training.

FACTOR 4 - (This factor was named BACKUP as the highest loading items address the issue of backing up the data.).

S3. Joe did not back up the data when he was supposed to.

S17. Joe has the responsibility for saving data at the proper

time.

S7. Joe gets along well with his co-workers.

S14. Joe is not required to backup the data as part of his job  
(reverse scaled).

### Choosing the Dependent Variable for Analysis

Out of the four scales constructed, using factor analysis as a guide, DEC-FAIR/PERFORM is the most suitable for measuring variance in evaluations of the target person. The other evaluation scale, GENERAL, unlike in the pilot study where it contained both general and specific items, with one exception contains all general items. Because general items are more prone to be biased, as discussed in the methodology chapter, DEC-FAIR/PERFORM is considered a superior measure of evaluation pertaining to the target person.

DEC-FAIR/PERFORM has the highest number of specific evaluative items relating to different aspects of work-related behavior of the target person loading on to it. Out of the five highest loading items on this factor, two are related to whether the supervisor's decision to fire the target person was fair and three assess the perceived performance of the target person. These items are clearly central to measuring the evaluation of the target person in the workplace context.

Furthermore, this factor-based scale has the highest reliability (.90) compared to the other scales. It, more than

other scales, can measure with precision the extent of favorability with regards to the evaluation of the target person. Based on these considerations, the hypotheses related to the grievant scenario were tested using DEC-FAIR/PERFORM as the dependent variable.

For the sake of giving a complete picture of the data, first some descriptive information will be provided relating to the independent variables and the various factors, including the OUTCOME variable. Note, however, that none of the other factor-based scales will be used to test the hypotheses. It will also become clear after examining the correlation matrix that a separate analysis on the OUTCOME variable, which is based on one item, cannot be justified in view of its fairly high correlation to the evaluation variable DEC-FAIR/PERFORM.

**Descriptive Statistics**

Descriptive information is provided on the number of observations at each class level of the independent variables.

**TABLE 7**

**THE GRIEVANT SCENARIO-UNIT DIRECTORS  
LEVELS OF THE INDEPENDENT VARIABLES**

<b>RATERSEX</b>	<b>FREQUENCY</b>	<b>PERCENT</b>	<b>CUM-FREQUENCY</b>
FEMALES	49	38.0	49
MALES	80	62.0	129

<b>TARGETSEX</b>	<b>FREQUENCY</b>	<b>PERCENT</b>	<b>CUM-FREQUENCY</b>
FEMALES	63	48.8	63
MALES	66	51.2	129

<b>GEN-ROLE</b>	<b>FREQUENCY</b>	<b>PERCENT</b>	<b>CUM-FREQUENCY</b>
NON-AGGRESSIVE	63	48.8	63
AGGRESSIVE	66	51.2	129

<b>SEQ-ORDER</b>	<b>FREQUENCY</b>	<b>PERCENT</b>	<b>CUM-FREQUENCY</b>
FIRST	70	54.3	70
SECOND	59	45.7	129

The means and standard deviations of the four factors as well as the OUTCOME variable are given below. They are followed by a correlation matrix containing the factors, the OUTCOME variable, and the independent variables.

**TABLE 8**

**THE GRIEVANT SCENARIO-UNIT DIRECTORS**

**THE MEAN AND THE STANDARD DEVIATIONS OF FACTORS AND OUTCOME**

VARIABLE	N	MEAN	STD DEV	ALPHA
GENERAL	129	6.10	1.0893	.88
DEC-FAIR/PERFORM	129	5.9968	1.2460	.90
TYPWORK/EFFORT	129	6.4853	1.2416	.86
BACKUP	129	7.9560	1.3735	.69
OUTCOME	126	3.3174	1.3425	N.A

**TABLE 9****THE GRIEVANT SCENARIO-UNIT DIRECTORS****CORRELATIONS OF FACTORS, OUTCOME, AND INDEPENDENT VARIABLES**

(THE LEVEL OF SIGNIFICANCE IS GIVEN UNDER THE CORRELATION)

	OUTCOME	GENERAL	DEC/PERF	BACKUP	TYPWORK
OUTCOME	1.00 0.0	-.3638 .0001	-.5651 .0001	-.0044 .9604	-.2592 .0034
GENERAL	-.3638 .0001	1.00 0.0	.5477 .0001	.0828 .3506	.4959 .0001
DEC/PERF	-.5651 .0001	.5477 .0001	1.00 0.0	.2342 .0075	.5098 .0001
BACKUP	-.0044 .9604	.0828 .3506	.2342 .0075	1.00 0.0	.4030 .0001
TYPWORK	-.2592 .0034	.4959 .0001	.5098 .0001	.4030 .0001	1.00 0.0
RATERSEX	.0029 .9742	-.0413 .6416	.0365 .6810	-.1399 .1138	-.0893 .3142
TARGETSEX	.0755 .4002	-.1221 .1680	-.1003 .2580	-.0115 .8968	-.1369 .1216
GEN-ROLE	-.0431 .6312	-.0859 .3331	-.0146 .8688	.0668 .4515	.0439 .6210
SEQ-ORDER	-.1359 .1292	.0424 .6333	-.0158 .8586	-.0842 .3425	-.0374 .6738

NOTE: DEC-FAIR/PERFORM IN THIS TABLE WAS ABBREVIATED TO DEC/PERF DUE TO LACK OF SPACE.

Note that in the scale measuring OUTCOME, low numbers indicate more favorable decisions and high numbers indicate less favorable decisions. Since in the factor-based scales, the higher numbers indicate more favorable evaluations of the grievant, while lower numbers indicate less favorable decisions, the correlations between OUTCOME and all the factor-based scales are negative.

It is apparent that OUTCOME is highly correlated to DEC-FAIR/PERFORM. This raises the question of whether OUTCOME can be usefully conceptualized as a separate variable revealing a different aspect of evaluation. This question is particularly pertinent to address since the OUTCOME variable is based on one item. Since, several items in the DEC-FAIR/PERFORM scale have about the same correlation to DEC-FAIR/PERFORM as does OUTCOME, analysis of OUTCOME as a separate variable cannot be justified. Therefore, the hypotheses will be tested only on the DEC-FAIR/PERFORM variable.

From the correlation matrix, it is clear that SEQ-ORDER and DEC-FAIR/PERFORM are not correlated. Since the sequence order appears to have had no effect on the evaluation of the target person as measured by DEC-FAIR/PERFORM, SEQ-ORDER will be dropped from further discussion in connection with the hypotheses pertaining to the grievant scenario.

It is also noteworthy that GEN-ROLE is not correlated to DEC-FAIR/PERFORM, raising the question of how well the gender-

role manipulation worked with the unit directors. This point, however, will be addressed in more detail in chapter 5 in the discussion section.

### Testing the Hypotheses

Next, information pertaining to the size of cells, as well as their means and standard deviations, is provided. Higher numbers indicate more favorable evaluations. Measurement was on a 9-point scale. Means, standard deviations and the n are given in each cell. Least square means are given at the bottom and are used to test the hypotheses. Note that least square means are very close in value to the unweighted means. This provides evidence that having an unbalanced design did not greatly impact the final results.



**TABLE 10****THE GRIEVANT SCENARIO-UNIT DIRECTORS****DEPENDENT VARIABLE: DEC-FAIR/PERFORM****RELIABILITY OF THE SCALE - .90****MEAN SQUARE ERROR = 1.50615297****MALE GRIEVANT****FEMALE GRIEVANTS****IN THE SCENARIO****IN THE SCENARIO**

<b>MALES</b> N=80	<b>AGGRESSIVE</b> MEAN=6.16 SD=1.52 N=25 LMEAN=6.16 CELL-A1	<b>NON-AGG</b> MEAN=5.82 SD=1.06 N=18 LMEAN=5.82 CELL-B1	<b>AGGRESSIVE</b> MEAN=6.33 SD=1.01 N=18 LMEAN=6.29 CELL-C1	<b>NON-AGG</b> MEAN=5.79 SD=1.27 N=19 LMEAN=5.73 CELL-D1
<b>FEMALES</b> N=49	<b>AGGRESSIVE</b> MEAN=4.86 SD=.51 N=10 LMEAN=4.83 CELL-A2	<b>NON-AGG</b> MEAN=6.20 SD=.924 N=13 LMEAN=6.20 CELL-B2	<b>AGGRESSIVE</b> MEAN=6.02 SD=1.43 N=13 LMEAN=6.02 CELL-C2	<b>NON-AGG</b> MEAN=6.43 SD=1.29 N=13 LMEAN=6.42 CELL-D2

The hypotheses involved a priori contrasts. T-tests were performed to test the hypotheses of no difference between the means. Both the null form and the alternate form of the hypotheses are given below along with the test results.

H01: Male and female managers will not differ in their evaluations of non-aggressive female grievants.

HA1: Male and female managers will differ significantly in their evaluations of non-aggressive female grievants, with males evaluating them higher than females.

Testing of this hypothesis requires that means of CELL-D1 and CELL-D2 be compared for difference.

Calculated T = -1.5504

p < .1238

RESULT: Unable to reject the null hypothesis at the .01 level of significance. The effect is in the direction opposite to the one predicted.

H02: Male and female managers will not differ in their evaluations of aggressive female grievants.

HA2: Male and female managers will differ significantly in their evaluations of aggressive female grievants with males evaluating them less favorably than females.

Testing of this hypothesis requires that means of CELL-C1

and CELL-C2 be compared for difference.

Calculated T =  $-.58914$

p <  $.5569$

RESULTS: Unable to reject the null hypothesis at the .01 level of significance.

HO3: Male managers will not differ in their evaluation of non-aggressive female grievants and male grievants.

HA3: Male managers will evaluate non-aggressive female grievants more favorably than non-aggressive male grievants.

Testing of this hypothesis requires that means of CELL-B1 and CELL-D1 be compared for difference.

Calculated T =  $-.21642$

p <  $.8291$

RESULTS: Unable to reject the null hypothesis at the .01 level of significance.

## RESULTS FROM THE MANAGER SCENARIO

A factor analysis with varimax rotation revealed three distinct factors embedded in the 50-item scale. The factor structure in this data set is similar to the one produced in the pilot study by the students. Using the .40 loading as the cutoff point, three different scales based on the factors were constructed. In the case of one item, S29, although it loaded at more than .40 on both the EFFORT and the LUCK/MERIT factors, it was decided to include it only in the LUCK/MERIT scale because it added more reliability to that scale. Having done that does not affect the hypotheses in any way since they are only tested on the evaluation scale GENERAL. Because of broad similarities, the names of the factors have been kept essentially the same as the names of the factors in the pilot study.

FACTOR 1 (GENERAL EVALUATION FACTOR)-High scores indicate more favorable evaluations of the target person.

S26. Ted is a fair and just manager.

S19. Ted is the kind of manager one can go speak openly with.

S39. Ted's personality is a great asset for him as a manager.

S37. Ted knows how to win the support of his subordinates.

S17. Ted knows how to get along with people.

S18. Ted values participative decision making.

S10. Ted does not treat his staff with respect (reverse

scale).

S5. Ted cares for the people who work for him.

G2. How would you rate Ted's interpersonal skills.

G4. How would you assess Ted's value as a team player.

S1. Ted acts fairly towards his staff.

S32. Ted has good interpersonal skills.

S13. Ted encourages his subordinates to express their views.

S11. Ted considers the opinions of his subordinates before making decisions.

S27. Ted knows how to effectively manage people.

S21. Ted can be described as a democratic leader.

S34. Ted has an effective management style.

S28. Working for Ted is probably quite stressful for his subordinates (reverse scale).

G8. How would you rate the contribution of Ted's personality to work place effectiveness.

S38. Ted's subordinates probably perceive Ted to be a competent leader.

S14. Ted's management style will hinder his future success (reverse scale).

S6. Ted is not handling his present job too well (reverse scale).

S15. Ted has been an effective manager at his present job.

S20. Ted does not have the personality to be a successful general manager (reverse scale).

- S7. Ted has the ability to be successful as a general manager.
- S3. Ted's sense of what is required for the health club is on the mark.
- S30. Problems facing Ted are due to factors outside his control.

FACTOR-2 (EFFORT)-High scores indicate a perception by the research subjects of a higher level of effort on the part of the target person.

- S22. Ted puts a lot of energy in his work.
- S25. Ted's job as the acting general manager is demanding.
- G7. How would you rate Ted's effort level.
- S16. Ted works hard at his job.
- G6. How would you rate Ted's level of work motivation.
- S35. The cause of Ted's career advancement lies in Ted himself.
- S8. Ted has a difficult task ahead of him.

FACTOR 3 (LUCK/MERIT)-High scores indicate that there is a perception of success due to factors internal to the target person. Low scores indicate perceived success due to chance or luck.

- S4. Ted has been pretty lucky in his promotions at the health club (reverse scale).
- S29. Ted's progress through the managerial ranks is no fluke.

S33. Ted's rise to the position of acting general manager has more to do with chance than with merit (reverse scale).

S40. Ted's past success was caused by factors outside of Ted (reverse scale).

### Choosing the Dependent Variable for Analysis

Out of the three scales constructed, using factor analysis as a guide, GENERAL is the most suitable for measuring variance in the evaluations of the target person. GENERAL taps the research subjects' attitude with regards to the management style and the effectiveness of the manager and has the highest number of specific evaluative items pertaining to managerial behavior loading onto it. These items are clearly relevant and central to measuring the evaluations of the target person (the health club manager) in the scenario.

Furthermore, GENERAL has a high reliability (.98). It can measure with precision the research subjects' evaluation of the target person. Based on these considerations, the hypotheses related to the manager scenario were tested using GENERAL as the dependent variable. For the sake of giving a complete picture of the data, however, first some descriptive information will be given relating to the independent variables and the various factors, including the OUTCOME variable.

Note, however, that none of the other factor-based scales

will be used to test the hypotheses. It will also become evident after examining the correlation matrix that the analysis of the OUTCOME variable based on one item can not be justified in light of its high correlation to the GENERAL.



**Descriptive Statistics**

Descriptive information is provided on the number of observations at each class level of the independent variables.

**TABLE 11**

**THE MANAGER SCENARIO-UNIT DIRECTORS  
LEVELS OF THE INDEPENDENT VARIABLES**

<b>RATERSEX</b>	<b>FREQUENCY</b>	<b>PERCENT</b>	<b>CUM-FREQUENCY</b>
FEMALES	49	38.6	49
MALES	78	61.4	127

<b>TARGETSEX</b>	<b>FREQUENCY</b>	<b>PERCENT</b>	<b>CUM-FREQUENCY</b>
FEMALES	64	50.4	64
MALES	63	49.6	127

<b>GEN-ROLE</b>	<b>FREQUENCY</b>	<b>PERCENT</b>	<b>CUM-FREQUENCY</b>
NON-AGGRESSIVE	62	48.8	62
AGGRESSIVE	65	51.2	127

<b>SEQ-ORDER</b>	<b>FREQUENCY</b>	<b>PERCENT</b>	<b>CUM-FREQUENCY</b>
FIRST	59	48.8	62
SECOND	68	51.2	127

The means and standard deviations of the four factors as well as the OUTCOME variable are given below. They are followed by a correlation matrix containing the factors, the OUTCOME variable, and the independent variables.

**TABLE 12**

**THE MANAGER SCENARIO-UNIT DIRECTORS**

**THE MEAN AND THE STANDARD DEVIATIONS OF FACTORS AND OUTCOME**

<b>VARIABLE</b>	<b>N</b>	<b>MEAN</b>	<b>STD DEV</b>	<b>ALPHA</b>
<b>GENERAL</b>	127	4.6420	2.2832	.98
<b>LUCK/MERIT</b>	127	6.0505	1.5609	.74
<b>EFFORT</b>	127	7.3175	1.1362	.79
<b>OUTCOME</b>	125	3.0160	1.6362	N.A

**TABLE 13**

**THE MANAGER SCENARIO-UNIT DIRECTORS**  
**CORRELATIONS OF FACTORS, OUTCOME AND INDEPENDENT VARIABLES**  
 (THE LEVEL OF SIGNIFICANCE IS GIVEN UNDER THE CORRELATION)

	<b>OUTCOME</b>	<b>GENERAL</b>	<b>LUCK/MERIT</b>	<b>EFFORT</b>
<b>OUTCOME</b>	1.00 0.0	-.5363 .0001	-.4182 .0001	-.2374 .0077
<b>GENERAL</b>	-.5363 .0001	1.00 0.0	.4021 .0001	.2701 .0021
<b>LUCK/MERIT</b>	-.4182 .0001	.4021 .0001	1.00 0.0	.3568 .0001
<b>EFFORT</b>	-.2374 .0077	.2701 .0021	.3568 .0001	1.00 0.0
<b>RATERSEX</b>	.0784 .3848	-.0002 .9981	.1254 1599	-.0171 .8482
<b>TARGETSEX</b>	.0002 .9979	-.0741 .4073	-.1006 .2604	-.1664 .0615
<b>GEN-ROLE</b>	.4712 .0001	-.8295 .0001	-.2544 .0039	-.1892 .0031
<b>SEQ-ORDER</b>	-.0205 .8198	.08759 .3275	-.0374 .6761	.0355 .6917

As in the grievant case, low numbers indicate more favorable outcomes and high numbers indicate less favorable outcomes. Since in the factor-based scales, the higher numbers indicate more favorable evaluations of the target person (the health club manager), while lower numbers indicate less favorable decisions, the correlations between OUTCOME and all the factor-based scales are negative.

The correlation matrix indicates that GENERAL is fairly highly correlated to OUTCOME at .54. The magnitude of this correlation is similar to what may be expected of an item actually belonging to the scale. This brings into question the idea of analyzing OUTCOME separately from GENERAL. Since the OUTCOME variable is based on one item, and since information based on single items tends to be unreliable, it is better to focus exclusively on GENERAL, which is a highly reliable scale measuring the evaluation of the target person. Therefore, the hypotheses were tested only with GENERAL.

From the correlation matrix, it is evident that SEQ-ORDER and GENERAL are not correlated. Since the sequence order appears to have had no effect on the evaluation of the target person as measured by GENERAL, SEQ-ORDER will be dropped from further discussion in connection with the hypotheses.

It is important to note that GEN-ROLE is strongly correlated to GENERAL. This provides evidence that the manipulation worked but also raises the question of whether

the manipulation was too strong. The implications of this, however, will be addressed in more detail in chapter 5 in the discussion section.

### Testing the Hypotheses

Next, information pertaining to the size of cells, as well as their means and standard deviations is provided. Measurement was on a 9-point scale. Higher numbers indicate more favorable evaluations. Means, standard deviations and the n are given in each cell. Least square means are given at the bottom and are used to test the hypotheses.

**TABLE 14****THE MANAGER SCENARIO-UNIT DIRECTORS****DEPENDENT VARIABLE-GENERAL****RELIABILITY OF THE SCALE - .98****MEAN SQUARE ERROR = 1.593554****MALE MANAGER****FEMALE MANAGER****IN THE SCENARIO****IN THE SCENARIO**

<b>MALES</b> N=78	<b>AUTOCRATIC</b> MEAN=2.77 SD=1.14 N=19 LMEAN=2.78 CELL-E1	<b>DEMOCRATIC</b> MEAN=5.99 SD=1.68 N=18 LMEAN=5.99 CELL-F1	<b>AUTOCRATIC</b> MEAN=2.88 SD=1.47 N=19 LMEAN=2.85 CELL-G1	<b>DEMOCRATIC</b> MEAN=6.68 SD=1.23 N=22 LMEAN=6.68 CELL-H1
<b>FEMALES</b> N=49	<b>AUTOCRATIC</b> MEAN=2.93 SD=1.02 N=14 LMEAN=2.96 CELL-E2	<b>DEMOCRATIC</b> MEAN=6.68 SD=1.34 N=12 LMEAN=6.93 CELL-F2	<b>AUTOCRATIC</b> MEAN=2.58 SD=.66 N=13 LMEAN=2.60 CELL-G2	<b>DEMOCRATIC</b> MEAN=7.27 SD=1.16 N=10 LMEAN=7.20 CELL-H2

The hypotheses involved a priori contrasts. T-tests were performed to test the hypotheses of no difference between the means. Both the null form and the alternate form of the hypotheses are given below along with the test results.

HO1: Male and female managers will not differ in their evaluations of autocratic female leaders.

HA1: Male and female managers will differ significantly in their evaluations of autocratic female leaders, with males evaluating them lower than females.

This hypothesis requires that the mean of CELL-G1 be compared with the mean of CELL-G2.

Calculated T =  $-.55715$

$p < .5785$

RESULTS: Unable to reject the null hypothesis at the .01 level of significance.

HO2: Male managers will not differ in their evaluation of autocratic female and male leaders.

HA2: Male managers will evaluate autocratic female leaders less favorably than autocratic male leaders.

This hypothesis requires that the mean of CELL-E1 be compared to the mean of CELL-G1.

Calculated T =  $-.1568$

$p < .8757$

RESULTS: Unable to reject the null hypothesis at the .01 level of significance.

### SUMMARY

This chapter started out with descriptive information about the demographic characteristics of the sample. Next, the response rate of the unit directors as well as the percentage of missing values in the two data sets were outlined. This was followed by results from the grievant scenario with regards to the factor analysis. Factor-based scales were outlined. The scale most appropriate for testing the hypotheses was then identified. Descriptive statistics on the complete data set were presented before presenting the results of the three hypotheses related to the grievant scenario. These steps were then repeated for the scenario involving the manager as a target person, and the results of the two hypotheses relating to that case were presented.

Of the 5 null hypotheses relating to evaluations of target persons in the grievant and the manager scenario, none could be rejected. The discussion section will focus on the explanations for the results.



## CHAPTER 5

### DISCUSSION OF RESULTS

#### INTRODUCTION

This chapter begins with an emphasis on the unique nature and features of this study which must be considered when interpreting the results. The results of the hypotheses testing with regards to the evaluation of the target persons in the manager and grievant cases are then discussed. The nature and quality of existing literature on gender bias is assessed and its role as a guide to future research is examined. The discussion concludes with both the strengths and limitations of the study and suggestions for future research.

#### Uniqueness of the Study

This study is unique for several reasons, and it is important to understand these reasons in order to put the results in their proper perspective. The following points should be considered in interpreting the results.

1. This is one of those rare field experiments in the gender bias literature where the subjects are managers. Typically, the overwhelming number of experimental studies use college students as subjects. The results of meta-analyses in the gender bias literature, as they relate to performance and leadership evaluation, such as those by Eagly et al. (1992)

and Swim et al. (1989), reflect this simple reality. Similarly, in the grievance resolution literature, true experiments with practitioners are rare.

2. Unlike other field experiments in the literature, particularly those conducted by mail, non-response of the subjects in this study was virtually non-existent, and the number of missing values was very low. This allows for stronger causal inferences which are not biased by methodological infirmities.

3. Both of the dependent variables used in the study (one for the case of the manager and one for the case of the grievant) have high reliabilities. Therefore, measurement error may be thought of as being minimal. With respect to using reliable measures, the present study with practitioners diverges from previous field experiments in the grievance resolution context which have generally used one-item measures or measures whose reliability is uncertain.

Next, the results from the grievant scenario as they relate to the hypotheses are discussed. This is followed by a discussion of hypotheses related to the manager scenario.

#### **DISCUSSION--THE GRIEVANT SCENARIO**

The null hypotheses related to the grievant scenario could not be rejected. Since all the hypotheses pertaining to the evaluation of the grievant stemmed from the same

conceptual framework, the results for the whole set are discussed together and their implications explored. The following four sections consider this issue in detail.

### The Effectiveness of the Gender Role Manipulation

The correlation matrix of independent and dependent variables from the grievant data set (in chapter 4, p. 171, Table 9) indicated that GEN-ROLE did not correlate highly with the evaluation variable DEC-FAIR/PERFORM. This raises an important issue with regards to whether the gender-role manipulation worked with the unit directors as it had in the pilot study with the students.

In order to examine more thoroughly the impact of GEN-ROLE on the responses of the unit directors, the General Linear Models (GLM) procedure using SAS was employed. This procedure is identical to the one used in the pilot study to explore the influence of the SEQ-ORDER and QUES-ORDER variables on the responses of student subjects as well as to check the GEN-ROLE manipulation and examine the psychometric properties of the pilot data.

The model for DEC-FAIR/PERFORMANCE was analyzed using RATERSEX, TARGETSEX, GEN-ROLE, and SEQ-ORDER as the independent variables. Because of unequal cell sizes, TYPE III sums of squares were used to derive results. The results of the analysis are given next.

Dependent Variable: DEC-FAIR/PERFORM (SS and MS have been rounded off to two decimal points.)

SOURCE	DF	TYPE III SS	MS	F Value	Pr >F
MODEL	15	28.5395	1.9	1.26	.2373
ERROR	113	170.195	1.51		
TOTAL	128	198.734			

R-Square=.143

C.V = 20.46

Root MSE = 1.227

SOURCE	DF	TYPE III SS	MS	F Value	Pr > F
RATERSEX(R)	1	.50	.50	.33	.5645
GEN-ROLE (G)	1	1.43	1.43	.95	.3321
TARGETSEX (T)	1	3.77	3.77	2.51	.1162
SEQ-ORDER (S)	1	.31	.31	.20	.6526
R*G	1	13.02	13.02	8.64	.0040
R*T	1	3.41	3.41	2.26	.1353
R*S	1	.03	.03	.02	.8812
G*T	1	2.55	2.55	1.69	.1962
S*G	1	2.84	2.84	1.89	.1724
S*T	1	.04	.04	.03	.8636
R*G*T	1	1.05	1.05	.70	.4055
R*S*G	1	2.85	2.85	1.89	.1718
R*S*T	1	.16	.16	.11	.7448
S*G*T	1	.70	.70	.46	.4971
R*S*G*T	1	.001	.001	.00	.9757

Confirming the results from the correlation matrix, GEN-ROLE does not enter into the model significantly. Although there is no main effect of GEN-ROLE as there was in the pilot study, there is an interaction between RATERSEX and GEN-ROLE ( $p < .004$ ). Clearly, the GEN-ROLE manipulation worked, but its effect was more subtle and unexpected. Next, the table of means for DEC-FAIR/PERFORM (Table 10) is replicated from page 174 and reexamined in light of the interaction between RATERSEX and GEN-ROLE.

## MALE GRIEVANT

## FEMALE GRIEVANT

## IN THE SCENARIO

## IN THE SCENARIO

MALES N=80	AGGRESSIVE MEAN=6.16 SD=1.52 N=25 LMEAN=6.16 CELL-A1	NON-AGG MEAN=5.82 SD=1.06 N=18 LMEAN=5.82 CELL-B1	AGGRESSIVE MEAN=6.33 SD=1.01 N=18 LMEAN=6.29 CELL-C1	NON-AGG MEAN=5.79 SD=1.27 N=19 LMEAN=5.73 CELL-D1
FEMALES N=49	AGGRESSIVE MEAN=4.86 SD=.51 N=10 LMEAN=4.83 CELL-A2	NON-AGG MEAN=6.20 SD=.924 N=13 LMEAN=6.20 CELL-B2	AGGRESSIVE MEAN=6.02 SD=1.43 N=13 LMEAN=6.02 CELL-C2	NON-AGG MEAN=6.43 SD=1.29 N=13 LMEAN=6.42 CELL-D2

The examination of the means reveals that the GEN-ROLE manipulation had a differential impact on male and female subjects. The female unit directors were apparently much more sensitive than the male unit directors to the aggressive behavior of the male grievants.

Female unit directors ranked non-aggressive grievants higher than aggressive grievants (regardless of the gender of the grievant) when compared to male unit directors, who showed just the opposite tendency. In particular, the female unit directors evaluated aggressive male grievants lower than aggressive female grievants ( $p < .026$ ). Female unit directors also evaluated aggressive male grievants lower than did the male unit directors ( $p < .005$ ).

So instead of finding a chivalry/paternalism effect on part of male managers when the target person was a female staying within her gender role and finding female managers, in contrast, to be unbiased in their evaluations, the results suggest a diametrically opposite conclusion. Indeed, what this study found can be described as the "nurturance/maternalism" effect (to coin a new term) on part of the female managers for the female grievants.

Male grievants, however, had no such luck. Female managers were, in fact, particularly critical in their evaluations of male grievants who were aggressive in the workplace context. This can be interpreted in a parallel

manner to the so called "evil women thesis" and be dubbed "the bad boy effect" (again coining a new term). Whereas the evil-woman thesis predicts that males in positions of power will give harsher evaluations and judgments to lower-level females who act out of their gender-role, the bad boy effect suggests that females in positions of power will more harshly evaluate lower-level males who continue to stay within the traditional male roles. These roles would include aggressive and assertive behavior in the workplace.

It is clear that GEN-ROLE as a manipulation did work in the grievant scenario. What is not clear is whether it would have worked in the same manner if it had been either more or less extreme. If we conceptualize gender-role behavior to lie on a continuum anchored on either side by two polar opposites (very non-aggressive and very aggressive), then it is reasonable to argue that the research subjects' sensitivity to it is going to differ depending on how far on the continuum (either to the left or the right) the target person lies.

It is also plausible to suggest, based on the interaction found in this study, that male and female managers have different thresholds at which the behavior of a target person becomes either favorably viewed or offensive. This study demonstrates that it is possible for the behavior of a target person, in this case an aggressive male subordinate, to be offensive to a female manager but not to a male manager.

Certainly, just the opposite scenario can be visualized, at least conceptually, where the behavior of a target person may be less favorably viewed by a male manager than a female manager. Indeed, there is evidence from the RATERSEX\*GEN-ROLE interaction that this can occur. For male managers, however, it was not the gender but the non-aggressive behavior of the target person that was less favorably evaluated.

Based on this line of reasoning, it can be argued that the null hypotheses related to the grievant case could not be rejected because the appropriate level of the GEN-ROLE manipulation needed to significantly influence the male managers was absent. This is of course just a possibility. In order to test such a possibility the GEN-ROLE manipulation would need to be varied at more than just two levels to find out how the level of the manipulation differentially affects the responses of the male and female managers. It is conceivable that GEN-ROLE manipulations of varying intensity would reveal other differences between male and female managers in their evaluations of male and female target persons in the grievance resolution context.

The finding in the present study, that the intensity of the GEN-ROLE manipulation may have a differential impact on male and female managers, leads to a more complex interpretation of the results. This is particularly true since the variance in responses of the female subjects due to



grievant gender does not fit neatly into the literature or some highly developed conceptual framework. The fact that the so-called chivalry/paternalism and the evil-woman theses only speak to the behavior of male decision makers points to a serious lack of symmetry in the grievance resolution literature. In light of the results in the present study, the necessity of inventing new terms as indicators for concepts, that may explain biased evaluations by female subjects, such as the nurturance/maternalism effect and the bad-boy effect, becomes obvious.

#### How the Results fit into the Literature

The results from the present study were not expected and have no precedent in the grievance resolution literature. However, one study in an entirely different context had at least one finding that was consistent with the present study. Scott (1983), using similar types of subjects (program agents reporting to county extension directors), found that the trust level of female respondents was significantly higher when they reported to a female manager rather than to a male manager.

Although Scott's study did not use the experimental design and manipulate gender-role, the results of the present study can be viewed in terms of the female subordinate/female manager trust relationship. Perhaps female unit directors did show partiality to lower-level female workers, regardless of

their gender-role behavior, because sameness in gender for females produces a mutually trusting relationship.

However, Scott's study was not in the grievance resolution context, and it only tapped the subordinates' trust perceptions towards the managers, and not vice versa. Therefore, whether managers reciprocate the trust of the same-sex subordinates by giving them higher evaluations and whether there is a difference in that reciprocity based on the gender of the manager is unclear based on empirical evidence. It could be argued on conceptual grounds that when managers give favorable evaluations to subordinates, it should be expected that subordinates would trust them more, although there may be differences with respect to the level of trust based on the respective gender of the manager and the subordinate.

But perhaps the causal arrow can go the other way. It could be argued that when subordinates trust their supervisors and this trust is manifested through behavior, the supervisors view the subordinates more favorably and that this attitude is reflected in evaluations and decisions pertaining to the subordinates. Again, the gender of both the superior and subordinate still may play a role in the evaluation of the subordinate in different contexts, including grievance resolution. However, isolating the effect of the gender of the subordinate requires the use of experimental methodology where the gender of the target person can be manipulated in a

laboratory setting.

Unfortunately there is not much literature on whether male or female managers would be influenced by the gender of the grievant in rendering their evaluations. The only two known experimental studies (Rosen & Jerdee, 1975; and Larwood et al., 1979) in the psychology literature that used managers as subjects to investigate biases in the grievance resolution context took place in the 1970s. Both of the studies, although widely cited in the literature pertaining to gender bias, were methodologically flawed.

Rosen and Jerdee (1975) used a two-item scale to form a dependent variable measuring the favorability of the evaluation of the grievance (and indirectly of the target person) by the research subjects. The authors did not indicate the reliability of their measure, and it is likely that the two-item scale had a low reliability. Furthermore, out of the 101 research subjects, only 23 were female managers. A low ratio of females to males in the sample makes it more difficult to make meaningful inferences with regards to differences based on sex of the subjects.

The field experiment by Larwood et al. (1979) was conducted by mail and had a response rate of 53%. Additional deletions were made in the sample due to incomplete responses and statistical considerations. For correlational analysis this effectively reduced the sample size to only 76 subjects

(see p. 542). The Larwood et al. study, in addition to suffering from non-response of subjects, which can substantially bias the findings of an experiment, used 8 dependent variables for analysis, each of which was based on one item. Using single items for analysis cannot, from a methodological perspective, be considered a sound practice and may lead to unstable results.

Both Rosen and Jerdee (1975), as well as Larwood et al. (1979), reported no differences between male and female managers as to how they evaluated male and female grievants. Both studies further concluded that female grievants in general were not as favorably evaluated as male grievants. The results of the present study examining gender bias in the grievance resolution context are inconsistent with the results of the only two published experimental studies using managers as subjects to examine work-related evaluations issues.

Even if the weaknesses in the methodology of the two earlier studies is ignored and it is assumed that the studies correctly concluded that the manager subjects (regardless of their gender) were biased against female grievants, it should not be expected that the results of 1970s can be replicated in the 1990s. As Gergen (1973) insightfully pointed out, the results in social psychological literature cannot be interpreted outside a historical context. Findings about how people evaluate and perceive something can change over a

period of time as values in a society change. Perhaps both the male and female managers of the 1990s are much different in how they view the workplace and the role of males and females than their counterparts in the 1970s. It is possible that male and female managers did not differ with regards to evaluations of male and female subordinates in the past, but they indeed do differ now.

Because of a changed historical context, how the male and female managers of today are different and how these differences are manifested in evaluations in the workplace context needs continual examination. It is possible that the change that has taken place in men (and therefore male managers) over the years due to the women's rights movement is of a different nature and quality than the change that has taken place in women (and female managers). It also may be that scholars have misinterpreted the magnitude, as well as the implications, of this change for the workplace. In this study we may be seeing the results of such a change, particularly as it pertains to evaluation judgments made by managers.

### **Questioning the "Chivalry/Paternalism" Effect**

The more recent experimental studies of the 1990s in the gender bias literature using arbitrators as subjects instead of managers differ in results from the Rosen and Jerdee and

the Larwood et al. studies of the 1970s in that they find evidence of a pro-female bias on part of male arbitrators. Unfortunately, they do not differ in terms of serious weaknesses in methodology. Bemmels' (1990) field experiment conducted through the mail, for example, suffered from a low response rate (the percentage of usable cases returned was 40.8%) and included 19 (14.5%) female arbitrators out of a total of 131. Oswald's (1991) study comparing the decisions of arbitrators and students included 146 arbitrators (the response rate of arbitrators was about 30%) in her sample, of which 29 were females. This type of a skewed male and female ratio combined with a high non-response ratio does not facilitate meaningful inferences with regards to how, and under what conditions, the gender of the subject and the gender of the target person interact and explain variance in evaluations judgments and decisions.

One wonders if the popularization of the so-called "chivalry/paternalism" effect in the arbitration literature is simply a function of the fact that the opportunity for the "nurturance/maternalism" effect is absent due to the low number of female arbitrators in the profession. However, without having the evaluations and decisions of the female arbitrators which can be compared to the evaluations and decisions of the male arbitrators, it is questionable whether the chivalry/paternalism effect can be meaningfully defined

and interpreted.

Even if it is accepted that male arbitrators do indeed show partiality to female grievants as opposed to male grievants, without knowing something about whether female arbitrators show the same type of partiality, one cannot come to conclusions about the nature of the effect and whether it has anything to do with the arbitrator being a male. Calling it the chivalry/paternalism effect becomes then a lame way to communicate that a sufficient number of female subjects (arbitrators) are simply not available for analysis. This is the current state of the literature and is reflected in the fact that many of the field studies that have found support for the "chivalry/paternalism" effect included no cases decided by female arbitrators (Bemmels, 1988a; 1988c). The studies that do include some female arbitrators contain very few cases decided by female arbitrators in which the worker grieving was also a female (Bemmels, 1988b, 1991).

In the present conceptualization of chivalry/paternalism, it is theoretically plausible and perfectly consistent for male arbitrators to render more favorable judgements to female grievants than male grievants and at the same time for these judgments to be overall much more or much less favorable when compared to the judgments of female arbitrators. The irony of course, from a research perspective, is that by defining the chivalry/paternalism effect in terms of the gender of the

target person, the gender of the research subject is automatically fixed. Then, for all practical purposes, the female research subject becomes irrelevant to the analysis.

Bemmels' (1991) field experiment conducted by mail with a response rate of 52% is an extreme example of this phenomena. In this study, the responses of 17 female arbitrators were eliminated because the number was too small for meaningful analysis and the focus of the study then shifted exclusively to the male arbitrators who had responded to written cases. What theoretical meaning can the finding of a chivalry/paternalism effect have in this case? Aside from the limitations imposed by the non-response of the subjects and the use of poor measures having uncertain reliability, the following question nags: How can we be certain that given a sufficient number of female arbitrators, if they could be included in the experiment, the results due to grievant gender would not be identical to the ones based on an analysis of the responses of male arbitrators only? Would we still call the effect a "chivalry/paternalism" effect?

Recently, an old notion known as the "queen-bee syndrome" (Staines, Tavis & Jayratne (1974) was revived and injected into the grievance resolution literature by Caudil and Oswald (1991) to provide some type of conceptual framework to include female arbitrators. This reflects the impoverished state of the theoretical literature in arbitration as the queen-bee



syndrome has failed to garner any empirical support (see Chapter 3) since it was first introduced in the management literature. It certainly did not find any support in the present study.

There have been many other field studies based on archival data both finding and not finding gender effects, and they are reviewed in Chapter 3. Despite the mixed evidence and a poor methodological base there is an emerging consensus that male arbitrators may be favorably biased towards female grievants, and it appears to have formed firm roots in the industrial relations literature. Arthur and Dworkin (1991), in their literature review, address the field studies by Bemmels (1988a, 1988b, 1988c). They state, "The results of all three studies were quite similar, although samples differed. There was substantial evidence that arbitrators tended to favor female grievants in determining whether to sustain a grievance against an unfair discharge" (p. 536). However, at least one major study based on archival data (consisting of two field assessments using large samples) conducted since the review did not find gender effects (see Dalton, Mesch, Owen, and Todor, 1992).

The fact that mixed evidence with regards to findings of the chivalry/paternalism effect in male arbitrators has led to some acceptance in the literature that such an effect actually exists is troubling. While future studies may (or may not)

find evidence to support the chivalry/paternalism thesis, the assumption in the literature (see for example, Arthur and Dworkin, 1991) that such an effect actually exists does not have overwhelming support. At best such a conclusion is premature given the rather serious methodological and measurement problems in the experimental studies using arbitrators as subjects as well as the inconsistent findings in the field studies.

It is possible that the null hypotheses could not be rejected because the research hypotheses are based on a stream of literature which is conceptually and methodologically flawed. Perhaps the chivalry/paternalism effect simply does not exist among male arbitrators as postulated by some of the arbitration scholars and that male arbitrators are in fact unbiased in their decision making. This is certainly not mentioned as a statement of fact but only as one possible explanation for the results of the current study. Given the previous discussion about the weaknesses in the literature, this possibility must be seriously considered.

#### **Would the Chivalry/Paternalism Effect Differ Between Managers and Arbitrators?**

Putting aside the conceptual and methodological issues with respect to studies that have found evidence for the paternalism/chivalry thesis, let us for a moment assume that

male arbitrators are indeed susceptible to at least a slight chivalry/paternalism effect that leads them to render more favorable judgments to the female grievants. Can a logical inference be made that such an effect would be similar for male managers in the work context?

Indeed, the conceptual framework for the present study borrowed heavily from the arbitration literature to infer a chivalry/paternalism effect for the male managers just as the arbitration literature had borrowed heavily from the management and criminology literature. Is it proper to make inferences by conceptually linking the evaluation judgments of male arbitrators with male managers and hypothesizing their similarity? What do male arbitrators and male managers have in common beside that they are male and is being a male enough to be influenced by the chivalry/paternalism effect when it comes to evaluating females in various contexts? That possibility must be explored.

First, it must be pointed out that although it is ironic, the framework laid out by Bemmels (1988a, 1988b, 1988c) to explore gender effects in arbitration was not original and, in fact, was taken from the Dalton and Todor (1985a) study which investigated whether *managers* (in a unionized environment) were biased in favor of females grievants in rendering judgments. It was the Dalton et al. field study and the two similar studies by the same first author and others (discussed

in the literature review) investigating gender bias in favor of females by *managers* that gave an impetus to the research stream investigating gender bias in favor of females by arbitrators. Since three of Bemmels' studies (1988a, 1988b, 1988c) and one study of Dalton et al. (1985a) came to similar conclusions about arbitrator and managerial decisions in a unionized environment, it provided grounds for arguing that the decisions of managers and arbitrators may share some similarity with respect to being influenced by the gender of the grievant.

Although the gender bias investigations of the 1980s in the grievance resolution context involving both managers and arbitrators share the same origins with respect to methodology (all of them being field studies based on archival data) as well as a reliance on concepts from the criminology literature, they have diverged considerably in the 1990s. Several field experiments have been conducted using arbitrators as subjects, but not a single field experiment has been conducted using managers as subjects. That situation is remedied by the present study. Indeed, this study provides a crucial link between the grievance resolution studies using managers as subjects and those using arbitrators as subjects.

It is well known that the large majority of the grievances in unionized companies are evaluated and decided by managers and are never appealed to arbitrators (Kochan, 1980).

In a non-union environment, managers must play an even more prominent role in evaluation of and settling of grievances. Managers have to take many other factors into consideration in deciding grievances other than the contract or company policy, although these are certainly relevant and play an important role.

Sometimes it may simply be expedient to settle a large number of union grievances at once (Fossum, 1989), and other times the morale of workers may be taken into account when deciding whether a demand should be acceded to or not (Meyer and Cooke, 1989). Managers, like arbitrators, may be influenced by the work histories of the grievants and are likely to render more favorable decisions to grievants with good performance records (Klaas, 1989). Although managers must consider many more work-related factors in the evaluation of grievances than arbitrators, it is still reasonable to argue that they, like the arbitrators, will be influenced by conscious or unconscious biases having to do with personal characteristics of the grievant such as gender.

Managers and arbitrators can be considered similar in many respects having to do with education and socio-economic status. Both groups consist of educated professionals who should not be expected to differ on how they perceive the male and female roles in the workplace. It can be argued that the type of work arbitrators and managers do is different enough

that they develop differing evaluation and decision-making schemas with respect to workplace situations. While perfectly plausible, this line of reasoning does not explain why the evaluation schema of male arbitrators would contain a pro-female bias while that of male managers would not.

There is one demographic difference, however, that needs to be addressed. Arbitrators, on the average, are older than managers. Henneman and Sandaver (1983), for example, reported the average age of an arbitrator as being close to 60. It is possible that age is a significant explanatory variable which was not incorporated in the model of this study. In other words, perhaps manifestation of chivalry/paternalism by males in authority is a function of their age as well as their gender. In this line of reasoning, older males would be more prone to showing a pro-female bias than younger males. Since male arbitrators are on the average older than male managers, it could be argued that whereas the chivalry/paternalism effect would be present among male arbitrators, it would not necessarily show up in a sample of male managers.

This line of reasoning, however, cannot be supported empirically. It is important to note that the age of the arbitrator is not discussed or included as an explanatory variable in many of the field studies by Bemmels (1988a, 1988b, 1988c) that found gender bias in favor of female grievants. Including age as an independent variable does not

seem to alter the results in other studies that have examined arbitrator behavior and decision making. A study by Henneman and Sandaver (1983) using a rather large sample of arbitration cases (1754) found that, in general, the personal characteristics of the arbitrators, including age, explained very little variance in their decisions. When arbitrator experience was included in the model in an experimental study (Bemmels, 1990a) as a proxy for age, no significant differences were found due to it. In a field study on arbitrator characteristics, Bemmels (1990b) found that older arbitrators tended to slightly favor the grievant (regardless of gender), although the magnitude of the estimate was considered too small to be of importance (p. 186). Oswald (1991) included age of the arbitrator as an independent variable in her model. The age of the arbitrator and grievant gender, however, did not interact and enter the model significantly. Clearly in the empirical literature, support is absent for the arbitrator age as an important factor in explaining gender bias in favor of females.

Therefore, it cannot be concluded that borrowing from the arbitration literature with regards to the findings on gender bias to infer how managers would make decisions is improper, particularly since the two streams of literature share the same origins. It can be said, however, that in real workplace situations managers probably do differ from arbitrators in

rendering judgments with regard to grievants. However, there are no plausible reasons to suggest that these differences in evaluations would imply that managers and arbitrators would differentially judge lower-level workers based on their gender. Nevertheless, if there are fundamental differences between the two groups with regard to how the work roles of male and females are perceived, which are not immediately apparent, the results could differ substantially depending on whether managers or arbitrators are used as research subjects.

#### **Summary of the Discussion--Grievant Case**

Discussions with regards to why the null hypotheses could not be rejected are always complex in nature as there are often numerous possibilities that can be considered as plausible explanations. For the hypotheses related to the grievant case, it can be said with some confidence that the inability to reject the null hypotheses is probably not due to measurement error or methodological inefficiency. Since the study is methodologically sound, significant results obtained about female unit directors instead of male unit directors cannot be rejected out of hand. Indeed, the findings may be emphasizing a portion of the conceptual landscape which has not yet been painted in the gender bias literature.

The findings of this experiment, like any well done experiment, have a straight-forward interpretation. Female



unit directors evaluated aggressive male grievants lower than aggressive female grievants on performance and abilities. These differentials in evaluation of the grievants were due to the manipulations involving gender and the gender-role behavior of the grievant. Furthermore, the evaluation of the aggressive male grievants by male unit directors relative to female unit directors was considerably more favorable. This suggests that gender and gender-role manipulations had a differential impact on the evaluations of male and female unit directors. Aggressive male grievants were clearly not evaluated positively by female unit directors but were evaluated positively by male unit directors.

It should be borne in mind that as intriguing as these results are, they must be interpreted cautiously, particularly since they appear independently of any conceptual framework. There is always the possibility that these results simply appeared by chance with this group of managers and cannot be replicated with other groups. Nevertheless, the results are striking and merit further investigation.

#### DISCUSSION--THE MANAGER SCENARIO

For both of the research hypotheses related to the manager scenario, the null hypotheses could not be rejected. There are many different reasons for why negative findings may have occurred in this case. These will be discussed in detail

in the following three sections.

### The Strength of the GEN-ROLE Manipulation

Based on the results of the correlation matrix, it is clear that the GEN-ROLE manipulation was quite successful in the case of the manager since it correlated highly to the evaluation variable. The magnitude of the effect is, however, of some concern. It is quite conceivable that the strength of the manipulation was so great that the more subtle effects due to the gender of the target person were simply drowned out. This possibility requires a serious investigation.

In order to examine more thoroughly the impact of GEN-ROLE on the responses of the unit directors in the case of the manager, the General Linear Models (GLM) procedure using SAS was again employed. The model for GENERAL was analyzed using RATERSEX, TARGETSEX, GEN-ROLE, and SEQ-ORDER as the independent variables. Because of unequal cell sizes, TYPE III sums of squares were used to derive results. The results of the analysis are given below and then interpreted.

Dependent Variable: **GENERAL** (SS and MS have been rounded off to two decimal points)

SOURCE	DF	TYPE III SS	MS	F Value	Pr >F
MODEL	15	479.994	31.99	20.08	.0001
ERROR	111	176.884	1.593		
TOTAL	126	656.878			

R-Square=.730

C.V = 27.19

Root MSE = 1.262

SOURCE	DF	TYPE III SS	MS	F Value	Pr > F
RATERSEX (R)	1	3.08	3.08	1.93	.1672
GEN-ROLE (G)	1	398.73	398.73	250.22	.0001
TARGETSEX (T)	1	.71	.71	.45	.5050
SEQ-ORDER (S)	1	0.00	0.00	0.00	1.000
R*G	1	3.87	3.87	2.43	.1219
R*T	1	1.14	1.144	.72	.3986
R*S	1	.002	.002	0.00	.9721
G*T	1	2.56	2.56	1.61	.2073
S*G	1	3.94	3.94	2.47	.1188
S*T	1	.11	.11	.07	.7921
R*G*T	1	0.0002	0.0002	0.00	.9897
R*S*G	1	4.26	4.26	2.67	.1049
R*S*T	1	.56	.56	.35	.5528
S*G*T	1	3.09	3.09	1.94	.1660
R*S*G*T	1	.67	.67	.42	.5166

It is clear from the above results that the GEN-ROLE manipulation strongly influenced the responses of the subjects. In fact, it explains most of the variance in the model. It appears then that the GEN-ROLE manipulation indeed may have made it difficult for other effects related to the gender of the target person to manifest themselves at a significant level. The table of means (Table 14) pertaining to the case of the manager is replicated from page 187 and reexamined for further analysis.

## MALE MANAGER

## FEMALE MANAGER

## IN THE SCENARIO

## IN THE SCENARIO

MALES N=78	AUTOCRATIC	DEMOCRATIC	AUTOCRATIC	DEMOCRATIC
	MEAN=2.77	MEAN=5.99	MEAN=2.88	MEAN=6.68
	SD=1.14	SD=1.68	SD=1.47	SD=1.23
	N=19	N=18	N=19	N=22
	LMEAN=2.78	LMEAN=5.99	LMEAN=2.85	LMEAN=6.68
	CELL-E1	CELL-F1	CELL-G1	CELL-H1
FEMALES N=49	AUTOCRATIC	DEMOCRATIC	AUTOCRATIC	DEMOCRATIC
	MEAN=2.93	MEAN=6.68	MEAN=2.58	MEAN=7.27
	SD=1.02	SD=1.34	SD=.66	SD=1.16
	N=14	N=12	N=13	N=10
	LMEAN=2.96	LMEAN=6.93	LMEAN=2.60	LMEAN=7.20
	CELL-E2	CELL-F2	CELL-G2	CELL-H2

The fact that there are large differences between the means for autocratic and democratic managers shows that the preference for the democratic and participative leadership style cuts across female and male unit directors. There is, however, an interesting contrast between male and female unit directors as to how they evaluated democratic managers in general. Female unit directors appeared to prefer democratic

managers (regardless of their gender) more than the male managers ( $p < .056$ ). Furthermore, female unit directors gave democratic female managers higher ratings than the ratings given to democratic male managers by male unit directors ( $p < .018$ ).

Although, it could be argued that these results appear purely by chance, there is support from the literature that women in general prefer a more democratic style of leadership to an autocratic one. In their meta-analysis of studies in the leadership domain, Eagly and Johnson (1990) reported "women tended to adopt a more democratic or participative style and a less autocratic or directive style than did men" (p. 233). They found this to be true across both laboratory and organizational studies. In light of this meta-analysis, it makes sense that female unit directors evaluated the democratic female manager in the scenario higher relative to the evaluation given by male unit directors to democratic male managers in the scenario.

#### **The Amount of Information Given in the Case**

Another plausible explanation that may explain why the null hypotheses were not rejected is the amount of information given in the case. The literature on sex-role stereotypes suggests that such stereotypes come into play most strongly when the available information about the target person is

limited. The more individuating information that is available to the evaluator about the target persons, such as their performance, behavior, interpersonal skills, etc., the less we should expect other characteristics such as race and gender to influence judgement with regards to them.

For example, a meta-analysis by Tosi and Einbender (1985) showed that evaluations of target persons are less influenced by sex stereotypes to the extent that more information about them is available. Since in this vignette, extensive information was available to the unit directors about the target person, it is possible that sex stereotypes did not have a chance to bias the evaluation.

Indeed, several unit directors with whom the researcher spoke to at each of the sites suggested that the case of the manager was much more clear cut than the case of the grievant. It is documented in the literature that evaluations are most prone to be influenced by stereotypes when the nature of the situation is ambiguous. However, when the facts of the situation are apparent, and extensive and unambiguous information is available about the target person, stereotypes are much less likely to influence the evaluations by subjects.

Based on this line of reasoning, it can be argued that the null hypotheses could not be rejected because a great deal of information was available in the vignette about the target person and because the situation described was unambiguous.

Therefore, stereotypes simply did not play a role in the evaluation of the target persons.

### The Role of the Organizational Composition

It has been suggested in the literature that the social composition of the organization may influence how a minority group (by sex or race) is perceived by others and evaluated by persons in authority (see the literature reviewed in Chapter 3 on sex stereotyping discussing works of Gutek and Morasch, 1982; Heilman, 1983; Kanter, 1977; Terborg, 1977). According to Kanter's (1977) theory of proportions, stereotyping of individuals is most likely to take place in an organization when the group (by sex or race) to which the individual belongs is very small compared to the majority group. Kanter suggested specifically that gender stereotypes would play a strong role in influencing evaluations when number of women in the work setting were comparatively few and the social composition of the organization was skewed. Kanter described a work situation as "skewed" when 15% or less of the individuals working belonged to a minority group.

This skewed social composition of the group certainly does not occur in the Virginia Cooperative Extension since the proportion of female unit directors is approximately 40%. Furthermore, approximately half of the top managers are females. It is clear, based on the proportion of female and



male field agents, unit directors, and top managers, that the culture of VCE is not male-dominated. Hence, sex stereotypes should not strongly influence evaluations of females. Perhaps, it should not come as a surprise that the female target person in the case of the manager was not evaluated significantly differently by male and female unit directors.

#### Summary of the Discussion-Case of the Manager

The responses to the case of the manager suggested no significant differences in the evaluations of autocratic female and male leaders by male and female unit directors. Therefore, the null hypotheses could not be rejected. Although many reasons can be postulated for the inability to reject the null hypotheses, the discussion focused on three explanations.

Again, the point must be made that, just as in the case of the grievant, measurement error pertaining to the dependent variable must be considered minimal. Indeed, a reliability of .98 for any dependent variable is extremely high. Therefore, it can safely be claimed that measurement error is not clouding the results of the study. Furthermore, although there were two non-respondents for the case of the manager a response rate of 98.4% must still be considered high. The non-response of two individuals should not be expected to bias the results of the experiment.

The results of this experiment, though not leading to the

rejection of the proposed null hypotheses, do have a couple of findings consistent with the gender bias literature pertaining to evaluation of managerial styles. These findings imply that, despite a strong gender-role manipulation and a clear and extensive description of the situation in the case, sex-role stereotypes may still have indirectly influenced the responses of both the male and female unit directors.

Evidence of this is found in the fact that there was a meaningful difference between how female unit directors evaluated democratic female managers and how male unit directors evaluated democratic male managers. Female unit directors gave democratic female managers in the case significantly higher evaluations than the evaluations given to democratic male managers by male unit directors ( $p < .018$ ). This finding is, as mentioned earlier, consistent with the leadership literature as it pertains to gender bias. There is also evidence of an overall preference by female unit directors for the democratic and participative leadership style. Therefore, the possibility that male and female managers may respond differently to and evaluate differently the performance of the same-sex manager (target person) when that manager is engaging in a democratic leadership style must be considered.

## CONCLUSIONS AND IMPLICATIONS

This study consisted of two field experiments with the same group of subjects--namely the managers of the Virginia Cooperative Extension (VCE), a large majority of whom are unit directors with the rest being higher-level administrators. Both field experiments took place at the same time. The sequence in which the case was given was controlled for by the design of the experiment.

Findings indicated that the sequence in which the cases were given to the unit directors had no impact on their responses. The null hypotheses pertaining to the cases of the grievant and the manager were not rejected. The possible explanations for the results were given in earlier sections. This study makes a significant contribution to the gender bias literature pertaining to evaluation in both the grievance resolution and the leadership literature in two ways.

1. This field experiment, with almost a perfect response rate, had a better ratio of male and female practitioners than any other field experiment conducted in the past using either managers or arbitrators as subjects. Including a reasonable proportion of female managers in studies is crucial if conceptual frameworks in the area of gender bias are to develop further. This is particularly true in view of the fact that female managers have increased significantly in the work force and will continue to increase in the future.

2. Since the overwhelming majority of grievances are decided by managers and not by arbitrators (both in union and a non-union environments), this study contributes to the literature by bringing the focus back to the important role that managers play in evaluations of grievances and grievants. In doing this, the study endorses the conceptualization of Dalton and Todor (1985a), who investigated pro-female bias on part of male managers in a field study, and viewed it as being important and worthy of serious pursuit.

#### MAKING SENSE OF IT ALL

What do the results of this study with the unit directors mean? Since the null hypotheses were not rejected, but there were still interesting findings, the answer must be framed carefully with the appropriate level of caution and enthusiasm. Leaving statistics and hypotheses testing behind for the moment, what conceptual inferences can be made from the results of these two experiments?

The results from these two experiments suggest that there is more variance in the responses of female unit directors rather than the male unit directors due to the manipulations. This finding cuts across contexts and appears to be true for both the case of the manager and the case of the grievant. This is surprisingly consistent particularly since the two cases were completely different and the GEN-ROLE manipulation

was perceived to occur at a weak level in the grievant scenario and at a very high level in the case of the manager. Nevertheless, in both cases involving different situations, female unit directors relative to male unit directors, were in general more favorably disposed to either female target persons or non-aggressive/democratic work-related behavior or both. If it is assumed for the moment that the results of the experiment are not due to chance or error, then the following implications have to be considered.

#### Having a Female Boss versus a Male Boss

It is possible that female managers consider gender-role behavior and gender of the person as a more important input to making workplace judgments than male managers. In view of the growing number of female professionals in the workplace, it may be important to have an understanding of how females in positions of authority and power will react to subordinates based both on their gender as well as workplace behavior. This knowledge would appear to be particularly useful for male subordinates having a female manager supervising them.

The results of the present study suggest that female managers may react positively towards both male and female workers who are non-aggressive. However, the same female managers may evaluate and judge aggressive males in the workplace context much more harshly, while tolerating or even

encouraging similar behavior on part of female workers. Therefore, it is possible that male workers reporting to a female boss may have an extra burden imposed upon them to be more careful about exhibiting certain types of aggressive behaviors in the workplace.

The implication is that while female workers may have more flexibility with regards to their behavior when they are reporting to a female boss, male workers must act in a non-assertive and a non-aggressive way in order to be evaluated and judged positively by female managers. However, workers reporting to male bosses, regardless of their gender, can be expected to be reacted to about the same. It should nevertheless be noted, based on the RATERSEX\*GEN-ROLE interaction found in the study, that male managers may respond more positively to aggressive behavior on part of lower-level employees in the workplace regardless of gender. This may be particularly true when the lower-level employee is asserting a work related right such as in the grievance resolution context.

### Possible Consequences

As mentioned in the literature review, differential treatment of male and females workers involving performance evaluations, promotions, and grievance resolution may have both legal and psychological consequences. First, differential

treatment may lead to a decline in satisfaction with management among workers who perceive that they are unfairly treated because of their gender (regardless of whether these workers are males or females). This may have consequences for the overall morale of the group and may lead to more serious work-related problems.

In addition to psychological consequences for the workers, organizations may want to be alert to possible legal consequences. If indeed lower-level male workers are treated and evaluated more harshly relative to lower-level female workers by female managers in different employment related contexts, this has the potential to cause serious legal problems.

The literature review pointed out how the Supreme Court used evidence from the social psychological literature on the role sex stereotypes play to support its decision in the Price Waterhouse case. That decision relied to a large extent on the psychological literature generated in the 1970s and early 1980s. If the finding regarding how female managers view aggressive male workers versus aggressive female workers is not an anomaly and finds further support, then this new literature capturing the historical context of the 1990s should be expected to affect the formulation of court opinions in the future.

Indeed, if the somewhat unexpected findings pertaining to

gender bias in the grievance resolution context have any validity, then organizations have good reason to be concerned. From a legal perspective, discrimination by female managers against male workers because they engage in a particular type of work behavior (aggressive, assertive) is no different than that of male managers against female workers because of sex-role stereotypes. In the current legal environment, where reverse discrimination is being debated as a serious workplace issue, these findings, if supported in further studies, may prove to be relevant.

In recent years, the interest in organizations in training male managers about gender-related issues has gained much popularity. Diversity training has become fashionable as organizations train their managers to get ready for the work force of tomorrow, which in large proportions will consist of females and minorities. Although it is assumed that it is the males in the organization that need to grow in sensitivity to the issues facing women workers, it may also be true that female managers may require similar training in the future in relationship to male workers whom they may be supervising. If indeed female managers feel more uncomfortable with aggressive male subordinates relative to female subordinates and this greatly influences how they evaluate male subordinates with regard to employment opportunities, then organizations will need to address this issue. This should be of particular



concern since in the future more male workers will be reporting to female supervisors.

### Preferred Managerial Style

The findings of the study also have implications for the type of management style that organizations may want to encourage because it has important consequences for the workplace environment. Both male and female unit directors responded generally positively to the democratic manager in the scenario. However, female unit directors seemed to respond more positively to the democratic style than the male unit directors. This has implications for middle-level managers who may be evaluated by senior female or male managers. Indeed, mid-level managers should consider the following ideas and suggestions based on the results of this study.

a. Mid-level managers should adopt a more democratic and a participative management style. It may be responded to positively by subordinates and evaluated more favorably by upper-level management than an autocratic and directive style where decisions are aggressively imposed on subordinates from the top.

b. Adopting a more democratic and participative management style may be viewed very favorably by female managers to whom they are reporting.

c. While adoption of a democratic and participative

management style by middle-level male managers will be responded to positively by other male managers to whom they might be reporting, or working with, it may be a less important input into their evaluations than into those of females reporting to other females.

Again, the results imply that the male or female subordinate, whether he or she is a lower-level worker or a manager, may want to consider the gender of his or her boss in exhibiting various types of work-related behaviors. This is because the gender of the subordinate as well as the level of dominance and aggression displayed may lead to a differential in work-related evaluations depending on the gender of the evaluating manager.

It should be emphasized that even if evaluations are only slightly influenced by gender, their long-term consequences cannot be underestimated. As noted by Larwood et al. (1979), the effect of negative or positive evaluations can accumulate over time and have serious consequences for an individual. For this reason the nature of management evaluations and decision making needs to be continually examined, particularly in how it is influenced by factors not directly related to work.

The research results may also have implications for top managers and the type of management style that is viewed positively. Although it is true that all the unit directors responded more positively to the democratic manager, the

female unit directors had the most favorable reaction. It may be important for senior managers to consider that female workers and female mid-level managers may value a democratic and participative management style more highly than male subordinates or managers. To the extent that such a leadership style positively impacts and develops the female subordinates, it may be worthwhile adopting it.

#### LIMITATIONS OF THE STUDY

Every study, no matter how well done, has at least some limitations. These limitations are now discussed within the conceptual framework of validity. The justification of this approach lies in the logic that all possible limitations of a study, in one way or another, imply that the results may lack validity from some perspective. Cook and Campbell (1979) outlined four different kinds of validities that researchers need to consider. These are internal, external, statistical conclusion, and construct validities. The relations and conflicts among these different types of validities have been addressed in the statistical and methodological literature (Cook & Campbell, 1979; Judd & Kenny, 1981), and it has been noted that often a researcher must make tradeoffs between different types of validities (Pedhazur & Schmelkin, 1991).

The issue of priorities among different types of validities is somewhat controversial (see Pedhazur &

Schmelkin, 1991). Cook and Campbell (1979), while emphasizing the importance of other validities, suggest that both in applied research as well as basic research internal validity is crucial and should have top priority. The authors do note the importance of other validities and suggest that for applied researchers external validity should be considered as quite important after internal validity. "The priority ordering for many applied researchers is something like internal validity, external validity, construct validity of the effect, statistical conclusion validity and construct validity of the cause" (Cook & Campbell, p. 83; although for a slightly different view on the matter, see Judd & Kenny 1981; and Olson & Peter, 1984).

As pointed out in Chapter 3, the explicit goal of this study, which consisted of two true experiments, was to maintain a high degree of internal validity. This was accomplished through the selection of a relatively homogeneous sample of subjects, random assignment, and the appropriate manipulations. The results of the study, however, may be questioned from different validity perspectives. Next, the limitations of the study are viewed from the conceptual framework of validity outlined by Cook and Campbell (1979).

### Internal Validity Considerations

Internal validity of research results may be thought of

as the extent to which (or the confidence with which) causal inferences can be drawn about the effect of one variable on another. Pedhazur and Schmelkin (1991) note, "...Internal validity is a necessary, although not sufficient, condition for external validity. Clearly when internal validity of a study is in doubt, it makes little sense to inquire to what or across what, are its findings generalizable" (p. 229). There are many potential threats to the internal validity of a study (discussed at length by Cook & Campbell, 1979; Babbie, 1992; Judd, Smith & Kidder, 1991; and Pedhazur & Schmelkin, 1991). The majority of these threats were controlled for in these two experiments with the process of random assignment (which results in groups equivalent on all extraneous variables). However, the following possibilities should be considered in evaluating the internal validity of this study.

First, the study was conducted over a period of two months at five different sites in Virginia. In an ideal experiment, it is best to have all people present at one place at the same time. This ensures that all extraneous influences are the same for everyone. However, since the study took place over many different sites at different times, it may be argued that the conditions under which the subjects read the cases and answered the questions may not have been constant. It would be particularly difficult to interpret the results if the unit directors who were subjects of the field experiment

in an earlier time phase talked about the study to other unit directors in different regions of Virginia who were to be future subjects. Such communication among the subjects, which leads to diffusion of information about the study, can dilute the results to an unknown degree thereby undermining the internal validity of the study (Babbie, 1992; Pedhazur & Schmelkin, 1991).

Second, to the extent that influences having to do with instrumentation affected the responses of the subjects, the results of the study may be contaminated. For example, it has been noted in the literature that (even though instruments given to subjects may be identical) the researcher could become progressively more proficient at using the instruments during the course of the study and therefore may vary in administering them from one set of subjects to another. The differences in instrument administration may subtly affect the responses of the subjects. This can also compromise the internal validity of the study (Pedhazur & Schmelkin, 1991).

Third, sometimes events (historical or otherwise) that take place during the course of the study can influence the attitudes of the subjects and thus their responses to the instruments. Since the VCE is generally engaged in an effort to sensitize the unit directors to sexual harassment and gender-related issues, the possibility that such training during the course of the study may have influenced the

responses of the subjects at some of the sites must be considered. To the extent that the unit directors were sensitized to gender-related issues, either through VCE programs or some other public event during the course of the study, and that exposure in turn affected their responses to the scenarios, the internal validity of the study is threatened (Pedhazur & Schmelkin, 1991).

### External Validity Considerations

It has been noted in the literature that maximizing internal validity (by selecting a homogenous sample of participants and settings) often must take place at the expense of generalizability of the results (Cook & Campbell, 1979; Judd & Kenny, 1982; Pedhazur & Schmelkin, 1991). From a purely statistical perspective, external validity is a function of selecting a sample at random from the population of interest to which one wishes to generalize. However, in social sciences, it is often not possible to draw a probability sample from the desired population (see Babbie, 1992; and Kerlinger, 1986 for good discussions of probability sampling theory). Therefore, speculations regarding generalizability of results from a study must be based on developed theoretical frameworks as well as common sense.

Ultimately, replications of the study with different samples at different sites using different methods is critical

to making inferences about or imposing limitations on the external validity of the results (see, Brewer and Hunter, 1989; Cook & Campbell, 1979; Judd, Smith and Kidder, 1991; and Pedhazur & Schmelkin, 1991). Based on this logic, the following points are noted that may hinder the generalizability of the results of this study to the population of managers as a whole.

First, the results of this study may not be generalizable if a *treatments-settings interaction* occurred (Cook & Campbell, 1979; Pedhazur & Schmelkin, 1991). Setting here refers to the type of environment in which the study took place as well as the nature of the study itself which was experimental. To the extent that influences unique to the sites where the field experiment was conducted affected the responses of the subjects, the results of the study would not be generalizable.

For example, the present sample consisted of unit directors and administrators who have faculty appointments and hold mid-level or higher management positions. This sample of subjects, however, may not be much like managers in private industry, or even other managers in the public sector. It is, therefore, quite possible that their responses to the written vignettes would be different than a sample of managers not sharing those characteristics. Furthermore, the extent to which the southern geographical location may have influenced



the responses of the subjects, the results may not be generalizable to other unit directors working for the cooperative extension of different states.

Second, the external validity of a study can also be undermined by a *treatments-attributes interaction* (Cook & Campbell, 1979; Pedhazur & Schmelkin, 1991). It should be noted that VCE has been very concerned about gender-related topics such as sexual harassment and other forms of sex discrimination. It is possible that male unit directors in the extension have developed a heightened awareness of gender-related issues. This heightened awareness may have affected their responses, and the responses in turn may be colored by the desire to appear unbiased. However, since much of the diversity training is often aimed at males and not females in organizations, it is possible that the female unit directors did not feel any particular need to restrain themselves in responding the way they actually felt. To the extent that male and female unit directors were influenced differentially by an awareness of gender-related issues in the organization and this affected their responses, the results of the study will lack in generalizability.

Finally, some of the findings of the study, particularly with regards to female subjects evaluating lower-level workers grieving a decision, fell outside any developed conceptual framework. However, it is particularly important for results

from a non-probability sample to fit in with some theoretical framework in order for a meaningful inference of external validity (Judd, Smith, & Kidder, 1991). To the extent that the findings are not consistent with any existing theory and may have occurred by chance, the results would not be generalizable to other managers in other organizations.

### Construct Validity Considerations

It is important to look at both the construct validity of the cause as well as the construct validity of the effect in order to make meaningful inferences about the research results. In a single study, it is often not possible or practical to operationalize constructs in more than one way. However, multiple operationism is crucial for generating evidence of construct validity (Brewer & Hunter, 1989; Pedhazur & Schmelkin, 1991). Since this study used a single method, method bias in both operationalizing the gender-role behavior construct as well as in measurement of the evaluation of the target person must be considered as a potential limitation in interpreting the results of the study.

The method used to operationalize the constructs of interest involved written cases followed by questionnaires tapping the attitudes of the subjects about the target persons in the scenarios. It has been noted that "Vignettes are a combination of expressive and objective ideas and projective

methods....Constructed with imagination and ingenuity, they can be interesting to subjects, can measure complex variables, can be good approximations to realistic and psychological situations (Kerlinger, 1986, p. 476)." Although the methods used in the present study have wide acceptance and are frequently used, still the point that no methodology is perfect and each has its own set of limitations has merit and is applicable.

Written vignettes, no matter how realistically they portray a situation and how extensive the amount of information given is in the case, cannot capture the quality of day-to-day interaction at the workplace between managers and subordinates. Therefore, it can be argued that results from employing such methods cannot be used to make direct inferences about how people would be evaluated in an actual work environment. This is particularly relevant since evaluations made about real workers by real managers have real consequences whereas evaluations of a target person in a written case do not.

There is certainly some validity to that criticism. On the other hand, real managers or workers cannot be directly asked if they are biased because of gender or race or some other characteristic of a person and be expected to be absolutely forthright, particularly since some of biases may be unconscious. Therefore, methods which unobtrusively tap

such attitudes have a place.

Certainly other methods, such as interviewing male and female managers or conducting focus groups with male and female workers and managers, would be extremely useful for shedding more light on gender issues in the workplace. It is generally recognized in the literature that a *triangulation* strategy (using multiple methods to measure the same phenomena) is essential for making valid inferences because it combines the strengths of different methods that have non-overlapping weakness (Fiske & Campbell, 1959; Pedhazur & Schmelkin, 1991). Brewer and Hunter (1989) state, "The multimethod strategy is simple, but powerful. For if our various methods have weaknesses that are truly different, then their convergent findings may be accepted with far greater confidence than any single method's finding would warrant" (p. 17).

The results of the study may also be considered limited by the fact that the GEN-ROLE manipulation (operationalizing the construct of gender-role) could only be varied at two levels in each of the scenarios. It is quite possible that there are other levels of the manipulation that may have given different results. However, answering that question was beyond the scope of this study. Because there are practical constraints, such as cell size, and the number of subjects actually available, not every level of the variable that is

desired can be included in a study. For example, in the grievant scenario, it would have been useful to have manipulated the gender of the supervisor (Stanley) as well as supervisor's gender-role behavior, along with the other manipulations. This certainly would have made for even a richer study. However, in individual studies, the scope as well as number of variables must be limited by practical considerations.

### Statistical Conclusion Validity Considerations

Statistical conclusion validity concerns have to do with the validity of conclusions based on statistical tests of significance. Such a discussion usually takes place in the context of issues such as Type I and Type II errors, effect size, and power of the statistical test (Cook & Campbell, 1979). As one important component in determining the power of the statistical test is the sample size, it merits special attention.

Because this study employed almost the whole population of unit directors and higher-level administrators in Virginia, the sample size could not have gone much above the 129 subjects that actually participated in the experiment. Since the sample size was small, the issue of whether there was enough statistical power is relevant.

Statistical power is a function of the alpha level set,

the effect size, the sample size, and the statistical test to be employed (see Howell, 1992; Pedhazur & Schmelkin, 1991). Since neither the sample size nor the effect size could be affected, an attempt was made to build the power into the design itself (see Lipsey, 1990). This was done by making sure that the dependent variables measuring the evaluation of the target persons in both cases were highly reliable (by first conducting a pilot study prior to the actual study). Reliable measures can lead to an increase in statistical power without a concurrent increase in the sample size (Devellis, 1991; Lipsey, 1990). Furthermore, the use of randomized block design using the sex of the subjects as a block aided in the number of female subjects in the cells being roughly equal to each other. Both of these approaches were adopted to reduce the amount of error variance and make the statistical analysis more efficient. Yet despite that, it is clear that having a larger sample may have led to more significant findings.

Having more statistical power may have provided stronger evidence for what the study seems to indicate. For the grievance resolution context, increased statistical power could have strengthened the conclusion that female managers evaluate aggressive male grievants negatively compared to aggressive female grievants and that there is a difference between male managers and female managers in how favorably they evaluate aggressive male grievants. For the case of the

manager, an increase in statistical power could have strengthened the conclusion that female managers, overall, prefer the democratic style of management more than do male managers.

It should be noted, however, that small samples are more prone to unstable results than large samples. Unfortunately, researchers often cannot control the size of their samples. Having 129 practitioners as subjects in a field experiment is not outside the norm. Of course, replication of this study with other groups of managers would be the best way to demonstrate whether the results are indeed valid or not.

#### **SUGGESTIONS FOR FUTURE RESEARCH**

Since the literature on how managers in the grievance resolution context may evaluate workers based on their gender is sparse, it is recommended that more work be done in this area. The results in the present study were inconsistent with both the experimental studies conducted with managers in the 1970s as well as those conducted with arbitrators in the early 1990s. However, these other studies typically contained a small number of female subjects, suffered from high non-response and did not use reliable measures.

It is suggested that more studies be conducted using practitioners as subjects and that these studies be on-site to avoid the problem of high non-response. Unless the gender bias

literature includes findings from studies that are well done, it cannot be relied on as a guide to future research. Unfortunately, at least in the context of grievance resolution, the present state of gender bias literature must be considered highly inadequate.

There is a dire need to expand and broaden the conceptual framework to include the possibility that female subjects may respond in a biased fashion against male target persons. The emphasis on the "paternalism/chivalry" thesis and the "evil woman thesis" by Dalton et al. (1985a) in the context of grievance resolution by managers and by Bemmels (1988a) in the context of grievance resolution by arbitrators, ignores completely the role female superiors may play in the workplace. With an ever-increasing flow of females into professional and managerial positions in which they have responsibilities to evaluate both male and female subordinates, it seems obvious that the theoretical framework must evolve to accommodate this new reality.

Therefore, it is critical that future research focus on more conceptual development in the gender bias area. For this to take place, female superiors in positions to exercise discretionary power in organizations must be included in studies in large numbers as research subjects. In addition, new concepts, terms, and definitions must be developed which will give some reality to the possibility that variance in the



evaluation of male and female target persons may occur when the evaluator is either a male or a female. The present study suggests that new terms such as the "nurturance/maternalism" effect and "the bad boy effect" need to be coined to create some type of symmetry in the gender bias literature, where more recently only the "chivalry/paternalism" effect and the "evil woman" thesis have played a prominent part.

Future researchers may want to follow the example of conducting more than one experiment in the same study to give insight into how the evaluations of male and female target persons can vary across contexts. For example, it can be argued, based on the evidence of the present study, that female evaluators may tend to view target persons engaging in a non-aggressive, consensus building workplace behavior more favorably than male evaluators. Male subordinates, who act aggressively in the workplace may be viewed as more threatening by female managers than by male managers. This idea should certainly be tested again. In particular, future studies can improve on the present work by having two target persons (the manager and the subordinate) being evaluated simultaneously in a case. This may lead to richer insights into how the interaction of the superior-subordinate dyad is perceived and how the evaluation of the superior and the subordinate are influenced by their respective genders and the gender of the evaluator.

## Conclusion

Future research in the gender bias area needs to be based on a broader conceptual framework. It should focus on how evaluations of males and females may differ due to their gender and to the extent that their workplace behavior is viewed differentially by male and female evaluators.

Furthermore, research should focus on the role of the gender-role manipulation in producing differential responses in male and female evaluators. It will be necessary to vary the intensity of the manipulation in many different contexts with many different samples to come to clear conclusions about how male and female managers in organizations vary in their evaluations of male and female subordinates as well as male and female leaders.

## REFERENCES

- Abramson, P.E., Goldberg, P.A., Greenberg, J.H., Abramson, L.M. (1977). The Talking platypus phenomenon: Competency ratings as a function of sex and professional status. *Psychology of Women Quarterly*, 2, 114-124.
- Abramson, L.Y., Seligman, M.E.P., Teasdale, J. (1978). Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology*, 87, 49-74.
- Allport, G.W (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.
- Arthur, J.B., & Dworkin, J.B (1991). Currents topics in industrial and labor relations research and practice. *Journal of Management*, 17(3), 515-551.
- Babbie, E. (1992). *The practice of social research*. Sixth edition. Wadsworth Publishing Company: Belmont, California.
- Babbie, E. (1990). *Survey research methods*. Second edition. Wadsworth Publishing Company: Belmont, California.
- Baron, R. (1988). Attributions and organizational conflict: The mediating role of apparent sincerity. *Organizational Behavior and Human Decision Processes*, 41, 111-127.
- Bartol, K.M. & Butterfield, D.A. (1976). Sex effects in evaluating leaders. *Journal of Applied Psychology*, 61(4), 446-454.
- Basil, D.C., & Traver, E. (1972). *Women in management*. New York: Dunellen
- Basow, S.A. (1986). *Gender stereotypes: Traditions and alternatives*. Monterey, CA: Brooks/Cole.
- Bayes, M., & Newton, P.M. (1978). Women in authority: A sociopsychological analysis. *Journal of Applied Behavioral Science*, 14, 7-20.
- Bemmels, B. (1988a). The effects of grievant's gender on arbitrators' decisions. *Industrial and Labor Relations Review*, 41(2), 251-262.
- Bemmels, B. (1988b). Gender effects in discipline arbitration:

- Evidence from british columbia. *Academy of Management Journal*, 31(3), 699-706.
- Bemmels, B. (1988c). Gender effects in discharge arbitration. *Industrial and Labor Relations Review*, 42(1), 63-76.
- Bemmels, B. (1990a). The effect of grievant's gender and arbitrator characteristics on arbitrator decisions. *Labor Studies Journal*, Summer/1990, 49-61.
- Bemmels, B. (1990b). Gender effects in grievance arbitration. *Industrial Relations*, 29(3), 513-525.
- Bemmels, B. (1991a). Attribution theory and discipline arbitration. *Industrial and Labor Relation Review*, 44 (3), pp. 548-563.
- Bemmels, B. (1991b). Gender effects in grievance arbitration. *Industrial Relations*, 30(1), 150-161.
- Bergman, B.R. (1986). *The economic emergence of women*. New York: Basic Books.
- Bigoness, W.J. (1976). Effect of applicant's sex, race, and performance on employers performance ratings: Some additional findings. *Journal of Applied Psychology*, 61(1), 301-311.
- Bigoness, W.J., & Dubose, P.B. (1985). Effects of gender on arbitrators' decisions. *Academy of Management Journal*, 28(2), 485-491.
- Bowman, G.E., Wortney, B.N., & Greyser, H.R. (1965). Are woman executive people? *Harvard Business Review*, 43(4), 164-178.
- Brewer, J., & Hunter, A. (1989). *Multimethod research*: Sage Publications Inc. Newbury Park: California.
- Brief for amicus curiae American Psychological Association in support of respondent, *Price Waterhouse v. Ann B. Hopkins* (1988). Washington, D.C: American Psychological Association, Executive Office.
- Broverman, I.K., Vogel, S.R., Broverman, D.M., Clarkson, F.E., & Rosenkrantz, P.S. (1972). Sex role stereotypes: A current appraisal. *Journal of Social Issues*, 28 (2), 59-78.

- Brown, V. & Gies, F.L. (1984). Turning lead in to gold: Evaluation of men and women leaders and the alchemy of social consensus. *Journal of Personality and Social Psychology*, 46(4), 811-824.
- Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, 56, 81-105.
- Cannings, K. & Montmarquette, C. (1991). Managerial momentum: A simultaneous model of the career progress of male and female managers. *Industrial and Labor Relations Review*, 44(2), 212-228.
- Cash, T.F., Gillen, B, & Burns, D.S. (1977). Sexism and "Beautyism" in personnel consultant decision making. *Journal of Applied Psychology*, 1974, 29, 80-85.
- Caudil, S.B., & Oswald, S.L. (1991). Ordinal logit estimation of biases in arbitration decisions. *Fifty-First Academy of Management Meetings*, Paper presentation.
- Caudil, S.B., & Oswald, S.L. (1992). An alternative to Bemmel's method of investigating biases in arbitration. *Industrial and Labor Relations*, 45(4), 800-805.
- Chesney-Lind, M. (1987). Female offenders: Paternalism reexamined. In L.L. Crites & W.L. Hepperle (Eds.). *Women, the Courts, and Equality*, Newbury, CA: Sage Publications, 114-139.
- Comrey, A.L. (1973). *A first course in factor analysis*. New York: Academic Press.
- Comrey, A.L. (1988). Factor analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology*, 56, 754-761.
- Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation. Design & analysis issues for field settings*. Chicago: Rand McNally.
- Crew, B.K.. (1991). Sex differences in criminal sentencing. Chivalry or patriarchy? *Justice Quarterly*, 8, 59-82.
- Dalton, D.R., and Todor, W.D. (1985a). Gender and workplace justice: A field assessment. *Personnel Psychology*, 38, 133-147.

- Dalton, D.R., and Todor, W.D. (1985b). Composition of dyads as a factor in the outcomes of workplace justice. *Academy of Management Journal*, 28, 704-712.
- Dalton, D.R., Todor, W.D., & Owen, C.L. (1987). Sex effects in workplace justice outcomes. *Journal of Applied Psychology*, 72, 156-159.
- Dalton, D.R., Mesch, D.J., Owen, C.L., & Todor, W.D. (1992). The 'Iron Law' of paternalism in organization justice outcomes? Two field assessments. *Fifty-Second Meeting of the Academy of Management Meetings*. Abstract Published in the Proceedings.
- Daly, K. (1989). Neither conflict nor labeling nor paternalism will suffice: Intersection of race, ethnicity, gender, and family in criminal court decisions. *Crime and Delinquency*, 35, 136-168.
- Deaux, K., & Emswiller T. (1974). Explanations of successful performance on sex-linked tasks: What is skill for males is luck for the females. *Journal of Personality and Social Psychology*, 29, 80-85.
- Deaux, K., & Lewis, L.L. (1983). Assessment of gender stereotypes: Methodology and components. *Psychological Documents*, 13, 25 (Ms. No. 2583).
- Deaux, K., & Taynor, J. (1973). Evaluation of male and female ability: Bias works two ways. *Psychological Reports*, 32, 261-262.
- DeVellis, R.F. (1991). *Scale development: Theory and applications*. Sage Publications Incorporated: Newbury Park: California.
- Dobbins, G.H. (1985). Effects of gender on leaders' responses to poor performers: An attributional interpretation. *Academy of Management Journal*, 1985, 28(3), 587-598.
- Dorfman, P., & Stephan, W. (1984). The effects of group performance on cognitions, satisfaction, and behavior: A process model. *Journal of Management*, 10, 173-192.
- Eagly, A.H. (1987). *Sex differences in social behavior. A social role interpretation*. Hillsdale, NJ: Erlbaum.
- Eagly, A.H., & Johnson, B.T. (1990). Gender and leadership style: A meta-analysis. *Psychological Bulletin*, 108(2),

233-256.

- Eagly, A.H., & Kraus, S.J. (1991). Gender and the emergence of leaders: A meta-analysis. *Journal of Personality and Social Psychology*, 60(5), 685-710.
- Eagly, A.H., Makhijani, M.G., & Klonsky, B.G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin*, 111(1), 3-22.
- Edwards, A. (1989). Sex/Gender, sexism and criminal Justice: Some theoretical considerations. *International Journal of the Sociology of Law*, 17, 165-184.
- Edwards, A. (1984). *Women on trial*. Manchester University Press: Manchester.
- Elig, T., & Frieze, I. (1975). A multidimensional scheme for coding and interpreting perceiver causality for success and failure events: The CSPS. *Catalog of Selected Documents in Psychology*, 5, 313 (MS 1069).
- Elig, T. & Frieze, I. (1979). Measuring causal attributions for success and failure. *Journal of Personality and Social Psychology*, 37(4), 621-634.
- Elson, A. (1981). Decisional thinking: Chicago panel report. *Proceedings of the Thirty Third Annual Meeting of National Academy of Arbitrators*, 62-118.
- English, H., English, A. (1958). *A comprehensive dictionary of psychological and psychoanalytical Terms*. New York: Langmans, Green and Co.
- Etaugh, C., & Sanders, S. (1974). Evaluation of performance as a function of status and sex variables. *Journal of Social Psychology*, 94, 237-241.
- Feather, N.T., & Simon, J.G. (1975). Reactions to male and female success and failure in sex linked occupations: Impressions of personality, causal attribution, and perceived likelihood of difference consequences. *Journal of Personality and Social Psychology*, 1975, 31, 20-31.
- Feldman, J.M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology*, 66, 127-148.
- Fiske, S.T., Bersoff, D.N., Borgida, E., Deaux, K., Heilman,

- M.E. (1991). Social science research on trial: Use of sex stereotyping research in Price Waterhouse v. Hopkins. *American Psychologist*, 46 (10), 1049-1060.
- Foschi, M. (1992). Gender and double standards for competence. In Cecilia L. Ridgeway (Ed.), *Gender, Interaction, and Inequality* (pp. 181-207). Springer-Verlag New York Inc.
- Fossum, J.A. (1989). *Labor relations: Development, structure, process*. Boston: BPI/Irwin.
- Frieze, I.H. (1976). Causal attributions and information seeking to explain success and failure. *Journal of Research in Personality*, 10, 293-305.
- Gerhart, B. (1990). Gender differences in current and starting salaries: The role of performance, college major, and job title, *Industrial and Labor Relations Review*, 43(4), 418-433.
- Gergen, K.J. (1973). Social psychology as history. *Journal of Personality and Social Psychology*, 26, 309-320.
- Goldberg, P. (1968). Are Women prejudiced against women? *Transaction*, 5 (5), 28-30.
- Green, S.G., & Mitchell, T.R. (1979). Attributional processes of leader-member interactions. *Organizational Behavior and Human Performance*, 23, 429-458.
- Gutek, B.A., & Morasch, B. (1982). Sex-ratios, sex-role spillover, and sexual harrassment of women at work. *Journal of Social Issues*, 38(4), 55-74.
- 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.
- Harlan, A., & Weiss, C.L. (1982). Sex differences in factors affecting managerial career advancement. In P.A. Wallace (Ed.), *Women in the Work Place* (pp. 59-96). Boston, MA: Auburn House.
- Harman, H.H. (1967). *Modern factor analysis*. First edition. Chicago: Universtiy of Chicago Press.
- Harman, H.H. (1976). *Modern factor analysis*. Second edition. Chicago: University of Chicago Press.



- Heilman, M. (1983). Sex bias in work settings: The lack of fit model. In B. Staw & L. Cummings (Eds.), *Research in Organizational Behavior* (Vol. 5, pp. 269-298). Greenwich, CT: JAI Press.
- Heilman, M., Block, C.J., Martell, R.F., Simon, M.C. (1989). Has anything changed? Current characterizations of men, women, and managers. *Journal of Applied Psychology*, 74, 935-942.
- Heilman, M., Martell, R.F., & Simon, M.C. (1988). The vagaries of sex bias: Conditions regulating the undervaluation, equivaluation, and overvaluation of Female Job applicants. *Organization Behavior and Human Decision Processes*, 41, 98-110.
- Heneman III, H.G, & Sandver, M.H. (1983). Arbitrator's background and behavior. *Journal of Labor Research*, 4(2), 115-124.
- Hopkins v. Price Waterhouse*, 618 F. Supp. 1109 (D. D.C. 1985).
- Hopkins v. Price Waterhouse*, No. 84-3040, slip op. (D. D.C. May 14, 1990) (on remand).
- Howell, D.C. (1992). *Statistical methods for psychology*. PWS-KENT Publishing Company.
- Ilgen, D. & Klein, H. (1988). Organizational behavior. *Annual Review of Psychology*, 40, 327-351.
- Issac, M.B. (1981). Sex role stereotyping and the evaluation of the performance of women: Changing trends. *Psychology of Women quarterly*, 6, 187-195.
- Issac, S., & Michael, W.B. (1981). *Handbook in research and evaluation*. EdITS Publishers, San Diego, California 92107.
- Izraeli, D.N. (1983). Sex effects or structural effects? An empirical test of kanter's theory of proportions. *Social Forces*, 62, 153-165.
- Jacobson, M.B., & Koch, W. (1977). Women as leaders: Performance evaluation as a function of method of leader selection. *Organizational Behavior and Human Performance*, 20, 149-157.
- Johnson, C. (1992). Gender, formal authority, and leadership.

In Cecilia L. Ridgeway (Ed.), *Gender, Interaction, and Inequality* (pp. 29-49). Springer-Verlag New York Inc.

- Johnson, D.R., & Scheuble, L.K. (1991). Gender bias in the disposition of case referrals: The effects of time and locations. *Criminology*, 29(4), 677-699.
- Jones, E.A. (1981). Decisional thinking of judges and arbitrators as triers of facts. *Proceedings of the Thirty Third Annual Meeting of National Academy of Arbitrators*, 45-61.
- Judd, C.M., & Kenny, D.A. (1981) *Estimating the effects of social interventions*. New York: Cambridge University Press.
- Judd, C.M., Smith, E.R., & Kidder, L.H. (1991). *Research methods in social relations*. Holt, Rinehart and Winston, Inc: The Dryden Press.
- Kanter, R.M (1977). Some effects of proportions on group life: Skewed sex ratios and responses to token women. *American Journal of Sociology*, 82, 965-950.
- Katz, D., & Braly, K.W. (1933). Racial stereotypes of 100 college students. *Journal of Abnormal and Social Psychology*, 28, 280-290.
- Kent, R.L., & Martinko, M.J. (1992). The measurement of attributions in organizational research. *Fifty-Second Meeting of the Academy of Management Meetings*. Abstract Published in the Proceedings.
- Kerlinger, F.N. (1986). *Foundations of behavioral research* (3rd Edition). New York: Holt, Rinehart & Winston.
- Kim, J.O. & Mueller, C.W. (1978a). *Introduction to factor analysis: What it is and how to do it*. Sage Publications Incorporated: Newbury Park, CA.
- Kim, J.O. & Mueller, C.W. (1978b). *Factor analysis: Statistical methods and practical Issues*. Sage Publications Incorporated: Newbury Park, CA.
- Klass, Brian S. (1989). Managerial decision making about employee's grievances: The impact of the grievant's work history. *Personnel Psychology*, 42, 53-67.
- Klineburg, O. (1951). The Scientific study of national

stereotypes. *International Social Science Bulletin*, 3, 505-515.

- Kochan, T.A. (1980). *Labor management relations research priorities for the 1980s: Final report to the secretary of labor*. Washington, D.C: U.S. Department of Labor.
- Kruttschnitt, C. (1982). Respectable women and the law. *The Sociological Quarterly*, 23(2), 221-234.
- Kruttschnitt, C. (1984). Sex and criminal court dispositions: The unresolved controversy. *Research in Crime and Delinquency*, 21, 213-232.
- Kryger, B.R., & Shikiar, R. (1978). Sexual discrimination in the use of letters of recommendations: A case of reverse discrimination, *Journal of Applied Psychology*, 63, 309-314.
- Larwood, L., Rand, P., Hovanessian, D. (1979). Sex differences in response to simulated employee discipline cases. *Personnel Psychology*, 32, 539-550.
- Liden, R., & Mitchell, T. (1985). Reactions to feedback: The role of attributions. *Academy of Management Journal*, 28(2), 291-308.
- Lippmann, W. (1922). *Public opinion*. New York: Harcourt, Brace.
- Lipsey, M. (1990). *Design sensitivity: Statistical power for experimental research*. Sage Publications Inc. Newbury Park: California.
- Locke, E.A. (1986). *Generalizing from the laboratory to field settings*. Lexington, MA: Lexington Books.
- Locksley, A., Borgida, E., Brekke, N., & Hepburn C., (1980). Sex stereotypes and social judgement. *Journal of Personality and Social Psychology*, 39, 821-831.
- Lott, B. (1985). Devaluation of women's competence. *Journal of Social Issues*, 41, 43-60.
- Martin, C.L. (1986). A ratio measure of sex stereotyping. *Journal of Personality and Social Psychology*, 52, 489-499.
- Martinko, M. & Gardner, W. (1987). The leader-member

- attribution process. *Academy of Management Review*, 12(2), 235-249.
- Mayes, S. (1979). Women in positions of authority: A case study of changing sex-roles. *Signs*, 4, 556-568.
- McElroy, J. (1982). A typology of attribution leadership research. *Academy of Management Review*, 7(3), 413-417.
- Meyer, D., & Cooke, W.N. (1988). Economic and political factors in formal grievance resolution. *Industrial Relations*, 27: 318-335.
- Miller, R.G. (1966). *Simultaneous statistical inference*. New York: McGraw Hill.
- Moulds, E.F. (1980). Chivalry and paternalism: Disparities of treatment in the criminal justice system. In Susan K. Datesman & Frank R. Scarpitti, eds. *Women, Crime, and Justice*. New York: Oxford University Press, PP. 277-299.
- Nagel, H. & Hagan, J. (1983) Gender and crime: Offense patterns and criminal court sanctions. In Michael Tonry & Norval Morris, eds., *Crime and Justice: An Annual Review of Research*, Vol. 4, Chicago: University of Chicago Press, pp. 91-144.
- Nieva, V.F., & Gutek, B.A. (1980). Sex effects on evaluation. *Academy of Management Review*, 5, 267-276.
- Nieva, V.F., & Gutek, B.A. (1981). *Women and work: A psychological perspective*. New York: Praeger.
- Nunnally, J. (1978). *Psychometric theory*. New York: McGraw Hill.
- Olian, J.D., Schwab, D.P., & Haberfield, Y. (1988). The impact of applicant gender compared to qualifications on hiring recommendations: A meta-analysis of experimental studies. *Organizational Behavior and Human Decision Processes*, 41, 180-195.
- Olson, J.C., & Peter, J.P. (1984). External validity? In P.F. Anderson & M.J. Ryan (Eds.), *1984 AMA Winter Educator's Conference: Scientific Method in Marketing* (pp. 81-84). Chicago: American Marketing Association.
- Oswald, S. (1991). Students as arbitrators: An empirical

- study. *Industrial Relations*, 30, 286-293.
- Peck, T. (1978). When women evaluate women, nothing succeeds like success: The differential effects of status upon evaluations of male and female Professional Ability. *Sex Roles*, 4, 205-213.
- Pedhazur, E.J., & Schmelkin, L.P. (1991). *Measurement, design, and analysis: An integrated approach*. Lawrence Erlbaum Associates Inc.. Publishers.
- Phillips, J.S., & Lord, R.G. (1982). Schematic information processing and perceptions of leadership in problem-solving groups. *Journal of Applied Psychology*, 67, 486-492.
- Powell, G.N. (1988). *Women and men in management*. Newbury Park, CA. Sage.
- Powell, G.N. (1990). One More time: Do female and male managers differ? *Academy of Management Executive*, 4, 68-75.
- Powell, G.N., & Butterfield (1989). The "good manager": Does androgyny fare better in the 1980's? *Group and Organization Studies*, 14, 216-233.
- Price Waterhouse v. Hopkins*, 109 S. Ct. 1775 (1989).
- Ragins, B.R., & Sundstrom, E. (1989). Gender and power in organizations: A longitudinal perspective. *Psychological Bulletin*, 105 (1), pp. 902-910.
- Rasinsky, K., Crocker, J., & Hastie, R. (1985). Another look at sex stereotypes and social judgements: An analysis of social perceiver's use of subjective probabilities. *Journal of Personality and Social Psychology*, 49, 317-326.
- Reed, G.C. (1989). Employer's new burden of proof in discrimination Cases. *Employment Relations Today*, Summer Edition.
- Ridgeway, C.L. (1988). Gender differences in task groups: A status and legitimacy account. In M. Webster, Jr., & M. Foschi (Eds.), *Status Generalization* (pp. 188-206). Stanford, CA: Stanford University Press.
- Rodgers, R.C., & Helburn, I.B. (1985) The arbitrariness of

arbitrators' decisions. *Proceedings of the Thirty-Seventh Annual Meeting of the Industrial Relations Research Association*, PP. 234-241.

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

Rosen, B., & Jerdee, T.H. (1974). The influence of sex-role stereotypes on personnel decisions. *Journal of Applied Psychology*, 59, 9-14.

Rosen, B., & Jerdee, T.H. (1975). Effects of employee's sex and threatening versus pleading appeals on managerial evaluations of grievances. *Journal of Applied Psychology*, 60, 442-445.

Russell, D. (1982). The Causal dimension scale: A measure of how individuals perceive Causes. *Journal of Personality and Social Psychology*, 42(6), 1137-1145.

Russell, D., & McAuley, E. (1986). Causal attributions, causal dimensions, and affective reactions to success and failure. *Journal of Personality and Social Psychology*, 50(6), 1174-1185.

Russell, D., 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*, 52(6), 1248-1257.

Russo, N.F., Kelly, R.M., & Deacon, M. (1991). Gender and success-related attributions: Beyond individualistic conceptions of achievement. *Sex Roles*, 25(5/6), 331-348.

Sackett, P.R., Dubois, C.L., & Noe, A.W. (1991). Tokenism in performance evaluation. The effects of work group representation on male-female and white-black differences in performance ratings. *Journal of Applied Psychology*, 76(2), 263-267.

Searle, S.R., Speed, F.M., and Miliken, G.A. (1980). Population marginal means in the linear model: An alternative to least squares means. *The American Statistician*, 34(4), 216-221.

Searle, S.R. (1987). *Linear models for unbalanced data*. New York: John Wiley & Sons, Inc.

- Sellin, T. (1928). The Negro criminal: A statistical note. *The Annals of the American Academy of Political and Social Science*, 140, 52-64.
- Schein, E.H. (1973). The relationship between sex role stereotypes and requisite management characteristics. *Journal of Applied Psychology*, 57, 95-100.
- Scott, C., & Shadoan, E. (1989). The effect of gender on arbitrator decisions. *Journal of Labor Research*, 10, 429-436.
- Scott, K.D. (1983). Trust differences between men and women in superior-subordinate relationships. *Group & Organization Studies*, 8(3), 319-335.
- Shuman, H., & Presser, S. (1979). Open and closed questions. *American Sociological Review*, 44, 692-712.
- Snyder, M. (1984). When belief creates reality. In L. Berkowitz (Ed). *Advances in Experimental Social Psychology* (Volume, 18, pp. 247-305).
- South, S.J., Bonjean, C.M., Markham, W.T., & Corder, J. (1982). Social structure and intergroup interaction: Men and women of federal bureaucracy. *American Sociological Review*, 47, 587-599.
- Staines, G., Tavris, C., & Jayaratne, T. (1974). The queen bee syndrome. *Psychology Today*, 7, 55-60.
- Steffensmeir, D.J. (1980). Assessing the impact of the women's movement on sex-based differences in the handling of adult criminal defendants. *Crime and Delinquency*, 26, 344-357.
- Stroh, L.K, Brett, J.M., & Reilly, A.H (1992). All the right stuff: A comparison of female and male managers' career progression. *Journal of Applied Psychology*, 77(3), 251-260.
- Struthers, C.W., Colwill, N.L., & Perr, R.P. (1992). An attributional analysis of decision making in a personnel selection interview. *Journal of Applied Social Psychology*, 22(10), 801-818.
- Summers, R.J. (1991). The influence of affirmative action on perceptions of beneficiary's qualifications. *Journal of Applied Social Psychology*, 21, 1265-1276.

- Swim, J., Borgida, E., Maruyama, G., & Myers, D.G. (1989). John McKay v. Joan McKay: Do gender stereotypes bias evaluation, *Psychological Bulletin*, 105, 409-429.
- Tabachnick, B.G., & Fidell, L.S. (1989). *Using multivariate statistics*. New York: Harper and Row.
- Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Social Issues*, 25, 79-97.
- Taynor, J., & Deaux, K. (1973). When women are more deserving than men. Equity, attributions, and perceived sex differences. *Journal of Personality and Social Psychology*, 32, 381-390.
- Teas, R., & McElroy, J. (1986). Causal attributions and expectancy estimates: A framework for understanding the dynamics of sales force motivation. *Journal of Marketing*, 50, 75-86.
- Terborg, J.R. (1977). Women in management: A research review. *Journal of Applied Psychology*, 62, 647-664.
- Terpstra, D.E., & D.D. Baker (1988). Outcome of sexual harassment charges. *Academy of Management Journal*, Vol. 31(1), pp. 185-194.
- Terpstra, D.E., & D.D. Baker (1992) Outcome of federal court decisions on sexual harassment. *Academy of Management Journal*, Vol. 35(1), 181-190.
- Thomas, K., & Tymon, W. 1982. Necessary properties of relevant research: Lessons from recent criticisms of the organizational sciences. *Academy of Management Review*, 7: 345-352.
- Top, T.J. (1991). Sex bias in the evaluation of performance in the scientific, artistic, and literary professions: A review. *Sex Roles*, Vol. 24 (1/2), pp. 73-106.
- Tosi, H.L., & Einbender, S.W. (1985). The effects of the type and amount of information in sex discrimination research: A meta-analysis. *Academy of Management Journal*, 28, 712-723.
- Tourangeau, R., & Rasinsky, K.A. (1988). Cognitive processes underlying contexts effects in attitude measurement. *Psychological Bulletin*, 1988, 103, 299-314.



- U.S. Bureau of the Census (1987). Household and family characteristics: March, 1987. In current *Population Reports* (Population Characteristics Series P-20, No. 424). Washington, D.C: U.S. Government Printing Office.
- Wall Street Journal*, September 25, 1992: A10.
- Wallston, B., & O'Leary, V. (1981). Sex makes a difference: Differential perceptions of men and women. *Review of Personality and Social Psychology*, 2, 9-41.
- Weber, R., & Crocker, J. (1983). Cognitive processes in the revision of stereotypic beliefs. *Journal of Personality and Social Psychology*, 45, 961-977.
- Weiner, B. (1974). Achievement motivation as conceptualized by an attribution theorist. In B. Weiner (Ed.), *Achievement Motivation and Attribution Theory*. Morristown, N.J: General Learning Press.
- Weiner, B. (1979). A Theory of motivation for some class room experiences. *Journal of Education Psychology*, 1979, 71, 3-25.
- Weiner, B. (1983). Some Methodological pitfalls in attributional research. *Journal of Education Psychology*, 75(4), 530-543.
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92(4), 548-573.
- Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer-Verlag.
- Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R.M. (1971). *Perceiving the causes of success and failure*. Morristown, N.J.: General Learning Press.
- Zingraff, Mathew and Randall Thomson (1984). Differential sentencing of women and men in the U.S.A. *International Journal of Sociology of Law*, 12(4), 401-414.
- Zwick, W.R., & Velicer, W.F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432-442.

## APPENDIX

THE APPENDIX CONTAINS THE INSTRUMENTS USED TO COLLECT DATA.

OUT OF THE EIGHT CASES GIVEN TO SUBJECTS, 4 ARE PROVIDED IN ORDER TO SAVE PAPER. THE OTHER FOUR CASES ARE IDENTICAL EXCEPT THAT THE TARGET PERSON IS OF THE OPPOSITE SEX HAVING A FEMALE NAME. BECAUSE OF LARGER MARGINS REQUIRED FOR DISSERTATIONS, THE CASES APPEAR TO BE SLIGHTLY LONGER IN PAGE LENGTH THAN THEY ACTUALLY ARE.

THE CASES PROVIDED ARE AS FOLLOWS

1. AUTOCRATIC MALE MANAGER (TED)  
(Case of the autocratic female manager named Kathy was identical and is not provided)
2. DEMOCRATIC FEMALE MANAGER (KATHY)  
(Case of the democratic male manager named Ted was identical and is not provided)
3. AGGRESSIVE FEMALE GRIEVANT (JANE)  
(Case of the aggressive male grievant named Joe was identical and is not provided)
4. NON-AGGRESSIVE MALE GRIEVANT (JOE)  
(Case of the non-aggressive female grievant named Jane was identical and is not provided)

EACH CASE IS FOLLOWED BY A QUESTIONNAIRE. THE SCALING INSTRUCTIONS AS WELL AS THE DEMOGRAPHIC QUESTIONNAIRE WHICH WERE PROVIDED TO THE SUBJECTS ARE INCLUDED AT THE END.

## THE CASE OF THE MANAGER

Ted, currently 33, has been the acting general manager for six months of an exclusive health and racquet club in New York City called the Elite Palace, which is part of a larger chain having 25 locations in the tri-state area. Ted has a reputation for being a determined person with an aggressive personality. Prior to becoming the acting general manager, Ted had been the marketing manager at Elite Palace for over 5 years. He is currently under review and is being considered for the position of permanent general manager by top management. Elite Palace is facing intense competition in the market place as many low cost clubs have been cropping up every where in Manhattan. The rate of growth in membership at the Elite Palace has slowed considerably over the last two years.

Ted was generally effective in meeting the goals for growth in membership put forth by corporate headquarters as the marketing manager for the first three years. The last two years have been somewhat frustrating for him as the health club market has become more saturated in New York City. People are demanding more services and think twice about paying the \$3500 dollar one time initiation fee to join Elite Palace (additional annual fee is \$1500). Since becoming the acting general manager, Ted has realized that the club has other problems as well. The health food restaurant at the club is barely breaking even. Other services that members have to pay extra for, such as masseurs and the masseuses, are being utilized much less frequently than in the past.

Corporate headquarters is located 5 blocks away from Elite Palace in mid-town Manhattan. Top managers frequently visit the health club and interact with the staff on an informal basis. These days the Elite Palace is full of rumors about Ted and his future at the club. Many people are convinced that Ted will become the next permanent general manager. However, there are others who believe that he is in for a big surprise. There is even speculation that Ted has been told to look for another job.

Elite Palace is a huge club, claiming the first 7 floors in a multi-storied building. It contains almost everything one could imagine a posh health club having. There are large swimming pools, saunas, tanning rooms, jacuzis, the latest nautilus machines to work out on. Exercise classes are ongoing at any given hour of the day from 7am to 10pm. These include classes in aerobic exercises, dancing, calisthenics, yoga, Tai Chi, karate, etc. A health food restaurant is located on the fourth floor to serve the health club clients, who sometimes eat there, or socially meet and drink with friends after working out. The restaurant employs 4 cooks and 7 part-time waitresses.

There is a sales staff of 18 people at the Elite Palace. The maintenance department houses 15 people to keep everything running smoothly. Suggestions regarding future

purchases of equipment come from this department, and are acted upon by purchasing agents. The health club employs 65 instructors who teach various types of exercise classes. Thirty of the instructors are full time, and 35 work on a part-time basis. These instructors report to floor managers (8 people). Floor managers are responsible for making sure that all classes get taught as scheduled, and that new members are trained adequately on the nautilus exercise machines so they do not injure themselves. Additional staff work as life guards, record clerks, and secretaries. Three assistant managers and the maintenance manager oversee the everyday operations of the club, and report to Ted.

Ted, although he is relatively young at 33, has been in the health club business for a long time. He started out as a part-time health club instructor at another club called the Viking, teaching calisthenics when he was in his first year of college. After a year, realizing that there was more money in sales, Ted started to sell health club memberships on a part-time basis at the Viking health club. Six months in sales convinced him to become a full-time sales person for the Viking health club. He continued in school part-time, and eventually received a Bachelors degree in business.

After three and a half years at Viking Health Club, Ted applied for the job of marketing manager at the Elite Palace Health and Racquet Club, which had just opened its third location in the heart of New York City. Ted did not get the job he applied for but was offered a position as sales representative at Elite Palace for higher base salary as well as commission. Ted accepted the job, and quickly proved himself to be a highly motivated and successful sales representative for the health club, opening up the most corporate accounts in his first year. He was promoted to the position of marketing manager after 3 years, when the old marketing manager left.

As the marketing manager Ted gained a reputation for being highly autocratic and directive with his staff. He is known for decisive action and does not tolerate dissent. Ted rarely seeks input from the staff before making important decisions. Once he has decided on a particular course of action, he expects everyone to fall in line. In staff meetings Ted does most of the talking, and is rarely questioned or contradicted by others.

Ted remained as the marketing manager for 5 years and was made the acting general manager when the previous general manager accepted a management position at a California health resort. Ted still continues to direct the sales staff. A sales manager will be hired, if Ted is approved by corporate headquarters as the permanent general manager of the Elite Palace.

When he was the marketing manager, Ted frequently accused the general manager of not focusing enough on advertising. His major gripe was that there was not enough support being given to the sales department. At management meetings he was often heard

to say, "Sales is our bread and butter at this club, and if new memberships dry up, we will all be history." It was rumored that the old general manager was fed up with some of staff, including Ted. The old general manager and Ted had clashed frequently over the last year before he left. In a management meeting, held almost a year ago, he sharply rebuked Ted when he thought Ted was getting out of line with his emphasis on advertising. "You think you know it all about this business, but I guarantee that you don't. Sometimes, I wish you would just shut up and listen. You may learn something. I bet that would be a surprise to you." Interestingly, when the general manager left without much notice, he recommended Ted as the interim manager to corporate headquarters.

Since becoming the acting general manager for the club, Ted has taken several steps to cut costs. He has laid off two of the masseurs and one masseuse with the least seniority. He has also reduced the hours of two cooks from 45 hours a week to 35 hours a week. He has put a moratorium on purchasing of new equipment until he has a chance to analyze the "true" needs of Elite Palace and its members. Ted has diverted some resources to marketing the Elite Palace in the media. This has generated lot of interest and the rate of visits by potential members to the club has doubled. However, the actual membership growth rate is about the same as it has been for the last two years, and that is fairly low.

Ted's exasperation became evident in a recent meeting when he lashed out at one of the sales people for bringing in a corporate client to the club during rush time in the evening. "We had the potential to bring in 150 new memberships. And you bring this guy in at 6 in the evening when the club is jam packed, and it looks like we have no more room. Damn it, how can you be so out of it. Listen to me carefully guys! Bring them in (potential members) between 8 to 11 in the morning or 2 to 4 in the afternoon, and the club will sell itself. It is virtually empty then." The sales person who Ted had addressed responded by saying that the guy wanted to come in at that time because he wanted to know what club was like in the evening when his guys would be coming in. Ted responded sharply. "You just don't get it. It is our job to sell the club. We have got to make sure these people come at good times, when they can see the club at its best. Once they are members we can deal with their complaints that the club is overcrowded at certain times. We can deal with everything afterwards, but first we've got to get them to sign up. It is really that simple! Damn it, do you need to be brain surgeons to figure this out! When some one tells you they want to come in after 5 in the evening, tell them we can better serve them in the morning, and really show them around. Tell them the advantages of giving their employees time off in the afternoon to work out. Talk to these people. They are not hard to convince. I know. I have done it a thousand times."

The maintenance and purchasing staff have been complaining that Ted has been

slow to approve purchase orders for bringing in new equipment that the previous general manager had agreed upon. Consequently, as membership increases, the older machines are being overworked, and frequently break down and have to be repaired. This upsets club members because during rush hours they have to frequently wait 4-5 minutes for a machine to be free before they can workout on it, and by that time they have often cooled down. Several floor managers have also indicated to the assistant managers that some of the current members are dissatisfied with the club, and are considering not renewing.

Ted has responded to this criticism by indicating his commitment to the present course of action. According to him, this is a transitional period for the club, and a change in strategy is necessary. His feeling is that although some of his decisions are not popular, in the long run they will help the club be more competitive. At a recent management meeting he said, "If somebody is not happy working here, they can just take a hike. We are going to have to buckle down and get the job done, otherwise none of us will be around too long in this place. My aim is to not let that happen. We are going to sell a lot of memberships, and we are going to have a lot of money coming in. We will get the best equipment out there, we will get more new machines. Right now we have to make do with what we have, which is also pretty good. Most other clubs can't even touch us with everything we offer to our members. If a machine breaks down during rush hour, we have got to fix it fast. That is why we have a large maintenance staff. I know we have problems during rush hour. We are working on that. This is going to take some time, but believe me, it is going to work. But if you don't support me, just get the hell out of this place. We don't need any whiners around here. They just bring everyone down."

Ted's potential promotion to general manager has generated lots of discussions among the staff, and is controversial. Ted feels that his record speaks for itself. The decision from top management is expected shortly. **Now put yourself** in the position of one of the senior managers at the corporate headquarters who is responsible for evaluating Ted. Indicate the extent of your agreement or disagreement with the following statements on a 9 point scale. **Nine indicates Strongly Agree** and **1 indicates Strongly Disagree**. Please remember that there are no right or wrong answers.

1. Ted acts fairly towards his staff.
2. Ted has been unsuccessful as the acting general manager.
3. Ted's sense of what is required for the health club is on the mark.
4. Ted has been pretty lucky in his promotions at the health club.

5. Ted cares for the people who work for him.
6. Ted is not handling his present job too well.
7. Ted has the ability to be successful as the general manager.
8. Ted has a difficult task ahead of him.
9. Ted is competent in what he does.
10. Ted does not treat his staff with respect.
11. Ted considers the opinions of his subordinates before making decisions.
12. Ted has a good performance record.
13. Ted encourages his subordinates to express their views.
14. Ted's management style will hinder his future success.
15. Ted has been an effective manager at his present job.
16. Ted works hard at his job.
17. Ted knows how to get along with people.
18. Ted values participative decision making.
19. Ted is the kind of manager one can go speak openly with.
20. Ted does not have the personality to be a successful general manager.
21. Ted can be described as a democratic leader.
22. Ted puts a lot of energy in his work.
23. Luck has played no role in the advancement of Ted's career.
24. Ted does not have the emotional maturity to be a general manager.
25. Ted's job as the acting general manager is demanding.

26. Ted is a fair and just manager.
27. Ted knows how to effectively manage people.
28. Working for Ted is probably quite stressful for his subordinates.
29. Ted's progress through the managerial ranks is no fluke.
30. Problems facing Ted are due to factors outside his control.
31. Ted successfully accomplishes his goals.
32. Ted has good interpersonal skills.
33. Ted's rise to the position of acting general manager has more to do with chance than with merit.
34. Ted has an effective management style.
35. The cause of Ted's career advancement lies in Ted himself.
36. Ted's leadership style is unlikely to change.
37. Ted knows how to win the support of his subordinates.
38. Ted's subordinates probably perceive Ted to be a competent leader.
39. Ted's personality is a great asset for him as a manager.
40. Ted's past success was caused by factors outside of Ted.

**Please give your ratings on a 9 point scale with 9 indicating the highest positive rating and 1 indicating the lowest negative rating.**

41. How would you rate Ted's overall performance.
42. How would you rate Ted's interpersonal skills.
43. How would you rate Ted's potential for future promotion.
44. How would you assess Ted's value as a team player.



45. How would you rate Ted's innate abilities and aptitudes.
46. How would you rate Ted's level of work motivation.
47. How would you rate Ted's effort level.
48. How would you rate the contribution of Ted's personality to work place effectiveness.
49. How would you rate Ted's overall leadership abilities.
50. How would you rate Ted as a manager.
- 51. Please pick the option you think is the most appropriate for top management to take. What is your recommendation.**
  1. Make Ted the permanent general manager, and give him a salary increase to go with the position.
  2. Continue to keep Ted as the acting general manager, give him a salary increase suitable for the position and monitor performance.
  3. Make Ted the marketing manager again, although at a higher salary than before, and hire another general manager.
  4. Make Ted the marketing manager again, with his salary at the same level, and hire another general manager.
  5. Transfer Ted to a smaller and less prestigious Elite Palace location as the marketing manager, with his salary staying at the same level.
  6. Transfer Ted to a smaller and less prestigious Elite Palace location as the marketing manager, with his salary reduced from the previous level.
  7. Give Ted one month's severance pay and fire him.

Please briefly note the reasons for your decision.

## **THE CASE OF THE MANAGER**

Kathy, currently 33, has been the acting general manager for six months of an exclusive health and racquet club in New York City called the Elite Palace, which is part of a larger chain having 25 locations in the tri-state area. Kathy has a reputation for being a caring person with a congenial personality. Prior to becoming the acting general manager, Kathy had been the marketing manager at Elite Palace for over 5 years. She is currently under review and is being considered for the position of permanent general manager by top management. Elite Palace is facing intense competition in the market place as many low cost clubs have been cropping up every where in Manhattan. The rate of growth in membership at the Elite Palace has slowed considerably over the last two years.

Kathy was generally effective in meeting the goals for growth in membership put forth by corporate headquarters as the marketing manager for the first three years. The last two years have been somewhat frustrating for her as the health club market has become more saturated in New York City. People are demanding more services and think twice about paying the \$3500 dollar one time initiation fee to join Elite Palace (additional annual fee is \$1500). Since becoming the acting general manager, Kathy has realized that the club has other problems as well. The health food restaurant at the club is barely breaking even. Other services that members have to pay extra for, such as masseurs and the masseuses, are being utilized much less frequently than in the past.

Corporate headquarters is located 5 blocks away from Elite Palace in mid-town Manhattan. Top managers frequently visit the health club and interact with the staff on an informal basis. These days the Elite Palace is full of rumors about Kathy and her future at the club. Many people are convinced that Kathy will become the next permanent general manager. However, there are others who believe that she is in for a big surprise. There is even speculation that Kathy has been told to look for another job.

Elite Palace is a huge club, claiming the first 7 floors in a multi-storied building. It contains almost everything one could imagine a posh health club having. There are large swimming pools, saunas, tanning rooms, jacuzis, the latest nautilus machines to work out on. Exercise classes are ongoing at any given hour of the day from 7am to 10pm. These include classes in aerobic exercises, dancing, calisthenics, yoga, Tai Chi, karate, etc. A health food restaurant is located on the fourth floor to serve the health club clients, who sometimes eat there, or socially meet and drink with friends after working out. The restaurant employs 4 cooks and 7 part-time waitresses.

There is a sales staff of 18 people at the Elite Palace. The maintenance department houses 15 people to keep everything running smoothly. Suggestions regarding future

purchases of equipment come from this department, and are acted upon by purchasing agents. The health club employs 65 instructors who teach various types of exercise classes. Thirty of the instructors are full time, and 35 work on a part-time basis. These instructors report to floor managers (8 people). Floor managers are responsible for making sure that all classes get taught as scheduled, and that new members are trained adequately on the nautilus exercise machines so they do not injure themselves. Additional staff work as life guards, record clerks, and secretaries. Three assistant managers and the maintenance manager oversee the everyday operations of the club, and report to Kathy.

Kathy, although she is relatively young at 33, has been in the health club business for a long time. She started out as a part-time health club instructor at another club called the Viking, teaching calisthenics when she was in her first year of college. After a year, realizing that there was more money in sales, Kathy started to sell health club memberships on a part-time basis at the Viking health club. Six months in sales convinced her to become a full-time sales person for the Viking health club. She continued in school part-time, and eventually received a Bachelors degree in business.

After three and a half years at Viking Health Club, Kathy applied for the job of marketing manager at the Elite Palace Health and Racquet Club, which had just opened its third location in the heart of New York City. Kathy did not get the job she applied for but was offered a position as sales representative at Elite Palace for higher base salary as well as commission. Kathy accepted the job, and quickly proved herself to be a highly motivated and successful sales representative for the health club, opening up the most corporate accounts in her first year. She was promoted to the position of marketing manager after 3 years, when the old marketing manager left.

As the marketing manager Kathy gained a reputation for being a kind and sympathetic, as well as a highly democratic manager. These days, a lot of the major decisions affecting the health club and its staff are generally made after a discussion of various alternatives and options. Kathy is the type of manager who likes to build consensus, and encourages the staff to ask questions openly and express their views.

Kathy remained as the marketing manager for 5 years and was made the acting general manager when the previous general manager accepted a management position at a California health resort. Kathy still continues to direct the sales staff. A sales manager will be hired, if Kathy is approved by corporate headquarters as the permanent general manager of the Elite Palace.

When she was the marketing manager, Kathy frequently suggested to the general manager that Elite Palace should focus more on advertising. Her major concern was that there was not enough support being given to the sales department. At management

meetings she was often heard to say, "Sales is our bread and butter at this club, and if new memberships dry up, we will all be history." It was rumored that the old general manager was fed up with some of staff, including Kathy. The old general manager and Kathy had clashed frequently over the last year before he left. In a management meeting, held almost a year ago, he sharply rebuked Kathy when he thought Kathy was getting out of line with her emphasis on advertising. "You think you know it all about this business, but I guarantee that you don't. Sometimes, I wish you would just shut up and listen. You may learn something. I bet that would be a surprise to you." Interestingly, when the general manager left without much notice, he recommended Kathy as the interim manager to corporate headquarters.

Since becoming the acting general manager for the club, Kathy has taken several steps to cut costs. She has laid off two of the masseurs and one masseuse with the least seniority. She has also reduced the hours of two cooks from 45 hours a week to 35 hours a week. She has put a moratorium on purchasing of new equipment until she has a chance to analyze the "true" needs of Elite Palace and its members. Kathy has diverted some resources to marketing the Elite Palace in the media. This has generated lot of interest and the rate of visits by potential members to the club has doubled. However, the actual membership growth rate is about the same as it has been for the last two years, and that is fairly low.

Kathy's exasperation became evident in a recent meeting when she disagreed with the approach of one of the sales people who had brought in a corporate client to the club during rush time in the evening. "We had the potential to bring in 150 new memberships. And you bring this guy in at 6 in the evening when the club is jam packed, and it looks like we have no more room. We can't afford to do this guys. Bring them in (potential members) between 8 to 11 in the morning or 2 to 4 in the afternoon, and the club will sell itself. It is virtually empty then." The sales person whom Kathy had addressed responded by saying that the guy wanted to come in at that time because he wanted to know what club was like in the evening when his guys would be coming in. Kathy nodded understandingly and said, "The main thing is that we have got to get them to sign up. Let us brainstorm on this. It is really important to bring them in at the right time. When some one tells you they want to come in after 5 in the evening, tell them we can better serve them in the morning, and really show them around. Tell them the advantages of giving their employees time off in the afternoon to work out. Talk to these people. They are not hard to convince. We need more members coming in during the morning and the afternoon. It is good for them and we can serve them better." Several other ideas about promoting the club were discussed at the meeting. There was active participation on the part of sales staff with Kathy listening carefully and making notes.

The maintenance and purchasing staff have been complaining that Kathy has been

slow to approve purchase orders for bringing in new equipment that the previous general manager had agreed upon. Consequently, as membership increases, the older machines are being overworked, and frequently break down and have to be repaired. This upsets club members because during rush hours they have to frequently wait 4-5 minutes for a machine to be free before they can workout on it, and by that time they have often cooled down. Several floor managers have also indicated to the assistant managers that some of the current members are dissatisfied with the club, and are considering not renewing.

Kathy has responded to this criticism by initiating several informal meetings every week to get more input from the staff. This format also gives Kathy a chance to explain why she thinks the present course of action is appropriate for the health club. According to her, this is a transitional period for the club, and a change in strategy is necessary. Her feeling is that although some of her decisions are not popular, in the long run they will help the club be more competitive. At a recent management meeting she said, "If somebody is not happy working here, and they have a legitimate gripe, I want to talk to them. My office is open. But people should realize that we are going to have to buckle down and get the job done, otherwise none of us will be around too long in this place. My aim is to not let that happen. We are going to sell a lot of memberships, and we are going to have a lot of money coming in. We will get the best equipment out there, we will get more new machines. Right now we have to make do with what we have, which is also pretty good. Most other clubs can't even touch us with everything we offer to our members. If a machine breaks down during rush hour, we have got to fix it fast. That is why we have a large maintenance staff. I know we have problems during rush hour. We are working on that. This is going to take some time but believe me, together we can make it work. I would really like for all of us to support each other in our jobs. This is a tough time for the club and I welcome your suggestions. This situation calls for teamwork."

Kathy's potential promotion to general manager has generated lots of discussions among the staff, and is controversial. Kathy feels that her record speaks for itself. The decision from top management is expected shortly. **Now put yourself** in the position of one of the senior managers at the corporate headquarters who is responsible for evaluating Kathy. Indicate the extent of your agreement or disagreement with the following statements on a 9 point scale. **Nine** indicates **Strongly Agree** and **1** indicates **Strongly Disagree**. Please remember that there are no right or wrong answers.

1. Kathy acts fairly towards her staff.
2. Kathy has been unsuccessful as the acting general manager.
3. Kathy's sense of what is required for the health club is on the mark.

4. Kathy has been pretty lucky in her promotions at the health club.
5. Kathy cares for the people who work for her.
6. Kathy is not handling her present job too well.
7. Kathy has the ability to be successful as the general manager.
8. Kathy has a difficult task ahead of her.
9. Kathy is competent in what she does.
10. Kathy does not treat her staff with respect.
11. Kathy considers the opinions of her subordinates before making decisions.
12. Kathy has a good performance record.
13. Kathy encourages her subordinates to express their views.
14. Kathy's management style will hinder her future success.
15. Kathy has been an effective manager at her present job.
16. Kathy works hard at her job.
17. Kathy knows how to get along with people.
18. Kathy values participative decision making.
19. Kathy is the kind of manager one can go speak openly with.
20. Kathy does not have the personality to be a successful general manager.
21. Kathy can be described as a democratic leader.
22. Kathy puts a lot of energy in her work.
23. Luck has played no role in the advancement of Kathy's career.
24. Kathy does not have the emotional maturity to be a general manager.

25. Kathy's job as the acting general manager is demanding.
26. Kathy is a fair and just manager.
27. Kathy knows how to effectively manage people.
28. Working for Kathy is probably quite stressful for her subordinates.
29. Kathy's progress through the managerial ranks is no fluke.
30. Problems facing Kathy are due to factors outside her control.
31. Kathy successfully accomplishes her goals.
32. Kathy has good interpersonal skills.
33. Kathy's rise to the position of acting general manager has more to do with chance than with merit.
34. Kathy has an effective management style.
35. The cause of Kathy's career advancement lies in Kathy herself.
36. Kathy's leadership style is unlikely to change.
37. Kathy knows how to win the support of her subordinates.
38. Kathy's subordinates probably perceive Kathy to be a competent leader.
39. Kathy's personality is a great asset for her as a manager.
40. Kathy's past success was caused by factors outside of Kathy.

**Please give your ratings on a 9 point scale with 9 indicating the highest positive rating and 1 indicating the lowest negative rating.**

41. How would you rate Kathy's overall performance.
42. How would you rate Kathy's interpersonal skills.

43. How would you rate Kathy's potential for future promotion.
  44. How would you assess Kathy's value as a team player.
  45. How would you rate Kathy's innate abilities and aptitudes.
  46. How would you rate Kathy's level of work motivation.
  47. How would you rate Kathy's effort level.
  48. How would you rate the contribution of Kathy's personality to work place effectiveness.
  49. How would you rate Kathy's overall leadership abilities.
  50. How would you rate Kathy as a manager.
  - 51. Please pick the option you think is the most appropriate for top management to take. What is your recommendation.**
1. Make Kathy the permanent general manager, and give her a salary increase to go with the position.
  2. Continue to keep Kathy as the acting general manager, give her a salary increase suitable for the position and monitor performance.
  3. Make Kathy the marketing manager again, although at a higher salary than before, and hire another general manager.
  4. Make Kathy the marketing manager again, with her salary at the same level, and hire another general manager.
  5. Transfer Kathy to a smaller and less prestigious Elite Palace location as the marketing manager, with her salary staying at the same level.
  6. Transfer Kathy to a smaller and less prestigious Elite Palace location as the marketing manager, with her salary reduced from the previous level.
  7. Give Kathy one month's severance pay and fire her.
- Please briefly note the reasons for your decision.



## THE CASE OF THE GRIEVANT

Jane, 24 years old and single, works in the claims processing department for the National Insurance company at one of its major regional offices. Jane is considered to be an energetic and aggressive individual with an independent personality. She has a two year technical degree in data processing, and joined the company about a year and a half ago after receiving her degree. Her job is to enter vital data about claims that are being processed or that have been settled into the computer. This includes statements made by all the involved parties, the claim adjuster's assessment of the situation, the potential liability associated with the claim, and sometimes the amount of money requested to obtain a release. There are approximately 15 other claims processors who work in the department. Two senior claims processors (Betty and Sam) assist the others, and train the new employees on various facets of the job if necessary. The department is run by the claims processing manager, who allocates the work and gives specific instructions.

The atmosphere in the department in the past generally has been informal and friendly. Disputes between supervisors and workers are pretty rare. The company has provisions for a formal appeal, if a worker is not satisfied with the decision of a supervisor. This policy is national in its scope and covers all offices and departments in various regions of the country. The appeals procedure involves the selection of a company ombudsperson who specializes in resolving grievances. The company ombudsperson is typically a manager who is not familiar with the grievant or the supervisor, and is completely neutral in the dispute. The decision of the ombudsperson is final as far as the company is concerned, and both the grievant and the supervisor must abide by it. National Insurance company is very proud of its ombudsperson program. It holds yearly training sessions that last for 2 weeks to train managers who want to serve in this capacity. The ombudspersons typically have around 15 years of seniority with the company along with reputations for fairness and impartiality in rendering decisions. Employees consider the management of National Insurance company to be very progressive, and feel they are treated fairly. Jane has been very happy with her experience at the National Insurance company until now. A series of incidents over the last 6 months have landed her into trouble. The facts of her situation follow.

Six months ago, upon the retirement of the previous manager Henry, a new manager of the claims processing department was hired. Stanley, the new manager is 45, has a B.A in finance, and has been working in the insurance industry for 23 years. He has had experience working as an insurance salesman as well as a claims adjuster.

The first thing Stanley did after being hired as a manager was to review prior performance of the people in the department using the previous manager's written general assessment of the staff. He held individual meetings with everyone in the department and

told them of his expectations as well as the goals they should shoot for. This created some initial stress in the department as the previous manager Henry only communicated about performance with the staff when there were serious problems, or during the routine performance review. Under Henry, performance review was usually a once a year, 15 minute informal conversation, going from topic to topic. Henry was always easy going and relaxed, and usually ended all his performance review sessions with comments like, "keep up the good work," "you are doing fine" etc.

Stanley, however, has a completely different style of management and the claims processing staff is attempting to adjust to it. Jane in particular became very concerned after her first meeting with Stanley. She doesn't know what others were told, but Stanley told her that her performance as judged by the previous manager Henry was marginal. Stanley said that according to the written general comments made by Henry to bring him up to date on the staff, Jane had a tendency to be easily distracted socially at the work place, and often had to be instructed several times before she adequately grasped a new task. Other information Stanley had obtained on productivity indicated that Jane's output, the number of claims processed per week, was in the bottom 10% of the staff. Stanley told Jane that her performance would be closely monitored, and that she should be making an effort to improve her performance.

Jane was surprised and responded sharply to Stanley's comments. She told Stanley that this was the first time she had heard this, and that Henry had never formally reviewed her. Jane further stated that Henry had only made positive comments to her about her work. She also asked Stanley whether he had taken into account the fact that a lot of the claims she handles tend to be very complex, involving serious accidents, and they take longer to process. Stanley said only that he was working with the raw data and the meeting was to identify general patterns and trends in the department. The meeting ended somewhat awkwardly, with Jane emphasizing fervently that not only numbers of claims but also the complexity of claims should be looked at for performance review purposes.

Two months later, Stanley called Jane in the office and talked with her again. "I have been told that you are taking 20-30 extra minutes on your lunch hour for the last week. What is going on?" "Nothing", replied Jane. "I have some friends who are visiting, and I have been eating with them. I make sure Betty or Sam (the senior claim processors) know that I am going to be late. I make up the time in the evening." "Well, you really need to focus on your work more. You mixed up information on two claims last week. I just got a call from the manager in the claims adjusting department, who told me that because of the mixup he has to rewrite both releases, which will delay the signing by at least a day. Sometimes, if you don't get those things signed quickly, people change their minds," said Stanley. "Listen Stan, I processed those claims exactly as I was given

by Betty. Paperwork for them probably got mixed up before they even got to me," responded Jane. "That may very well be, but that is the kind of thing we should be able to catch here. Don't you think so?" said Stanley as he pressed on. "That's ridiculous Stanley. I don't have time to catch everything around here. I know other people here who have made those kind of mistakes before. They do happen. I am sorry but I am not perfect. But I really resent your suggesting that I am not doing my job. You have done that before for no reason and you are doing it again. Just get off my back," said Jane as she got up and walked out. It was clear that relations between Stanley and Jane had become tense.

Three months later, another incident took place that has led to the current dilemma faced by Jane. All the claim processors, typically before leaving for the day, back up the data they have entered to ensure safety. This practice, although required by company policy is not always followed. In fact some claims processors have been known to back up data only once a week on Friday. Last Friday afternoon after lunch, Jane was asked by Sam, one of the senior claim processors, to assist in training a new data-entry person. The training took the whole afternoon, and since Jane was not at her terminal, she forgot to backup the data. She remembered when she got home and gave Sam a call at home to tell him what had happened. "Ah, hell don't worry about it. Do it on Monday morning." Jane went to work as usual on Monday, and started processing new claims. After a couple of hours she went to get some coffee, and when she came back to the terminal, the screen was stationary. Jane hit the keyboard and nothing happened. She realized that something was wrong and called Sam. Sam played with the computer and said that the hard disk had crashed and phoned the technician.

It soon became evident that Jane had lost some very important data, as she had not backed up her work on Friday or her work on Monday morning. Sam assured her it would be O.K. They would simply get the original paper work, and Jane could enter it all in again. When Stanley heard of this, however, he was furious and called a department meeting. At the meeting he said, "I think you all know how important claims data is to the functioning of this company. The people we insure trust us to process their claims and settle them as speedily as possible. In order to do that we have to take great care in collecting and maintaining this data. Frankly, this department, or I should say some people in this department, have developed bad habits regarding how claims data is to be processed and treated. This is a serious job, and we can not be frivolous about it. I want you all to read the company manual on claims processing, especially the section on the importance of backing up data. According to the company policy, an employee responsible for losing claims data can be summarily dismissed."

After the meeting ended, Stanley called Jane in his office. He said, "I am sorry Jane, but I really don't think you belong here. The kind of work we do really requires

an aptitude for and mastery of detail, and frankly you just don't have it. I think it will be in your interest to find something more suitable for yourself. I will be happy to give you a recommendation letter." Jane was completely stunned and speechless. After a few seconds she finally said, "Stanley, I don't know what you have against me but you are being completely unfair." Stanley, however, continued, "I have been thinking about this off and on. Your performance is poor in general in this department. I think you are going to be better off, and we are going to be better off." Jane got up from the chair and said, "Stanley, you are a jerk and a lousy manager and no body in this department can stand you. Its not me but you who is incompetent. You are the one who doesn't belong here Stan. You can't make this stick. You will end up looking like a moron." Stanley did not say anything. Jane's face was red and she was absolutely furious as she left. She saw Sam and Betty talking outside. She went up to them and said "That S.O.B Stanley just fired me." "Are you serious," said Sam. "Yes I am, and I am going to appeal this. He is not getting away with this," said Jane.

Sam and Betty, the two senior claim processors both went in to talk to Stanley. Stanley, however, would not budge. "Jane is not as bad as you think Stan. She gets her work done on time most of the time. I know she is a little aggressive sometimes, but she pitches in. She likes to train newcomers and show them the ropes. That helps us all out. That is what I was having her do when she forgot to backup. Stanley said that he had the discretion to fire workers when he felt their performance was below par. He told Sam and Betty that after observing Jane for 6 months, he felt he was making an informed decision. "She is going to appeal this," said Betty. "You are of course familiar with the appeals procedure at the company," asked Sam. "Look, both of you, I have made my decision. I run this department and I have made my decision. She can appeal it to whomever she wants but this is her last week in this department. I understand your sympathy for her, but my goal is to have a smooth functioning and an efficient department. This means that some people are going to have to start performing better than they have been. If they are unwilling to or can't, then they should find a more suitable position elsewhere."

Jane did appeal the decision. A senior manager, acting as the ombudsperson from a different location was selected to hear the grievance. Put yourself in the position of that manager and indicate the extent of your agreement or disagreement with the following statements on a 9 point scale. **Nine** indicates **Strongly Agree** and **1** indicates **Strongly Disagree**. Please remember that there are no right or wrong answers.

1. Stanley is a fair manager.
2. Jane has the ability to be a good performer.

3. Jane did not back up the data when she was supposed to.
4. Jane's mistake was a freak accident.
5. Jane is likely to make serious mistakes at the workplace in the future.
6. Jane's job is an easy one.
7. Jane gets along well with her co-workers.
8. Jane does not have an aptitude for detail work.
9. Jane contributes significantly to the overall department effectiveness.
10. The cases assigned to Jane are complex.
11. Jane's future performance is likely to be good.
12. Jane is careless at work.
13. Jane successfully accomplishes the tasks assigned to her.
14. Jane is not required to backup the data as part of her job.
15. Jane is a conscientious worker.
16. Jane does not come up to the standards of the department.
17. Jane has the responsibility for saving data at the proper time.
18. Jane has no one to blame but herself for her predicament.
19. Jane is easily distracted from her work.
20. The work performed in Jane's department requires specialized training.
21. Stanley's decision regarding Jane lacks basic fairness.
22. Jane's failure to backup the data was due to a factor outside her control.
23. Jane is competent in what she does.

24. Jane has good work habits.
25. Stanley's decision to fire Jane is an overreaction.
26. The type of work Jane performs is relatively routine.
27. Jane does not put enough effort in her work.
28. What happened to Jane could have happened to anybody.
29. Stanley made a well thought out decision when he fired Jane.
30. Jane is a highly skilled worker.
31. Bad Luck played a major role in Jane's predicament.
32. Jane works hard at her job.
33. Jane is helpful at the work place.
34. Jane has a demanding job.
35. Jane is a good performer at the work place.
36. Jane puts a lot of energy in her work.
37. Jane's mistake reveals something about Jane that is unlikely to change in the future.
38. Jane's present situation is a result of Stanley's unfairness.
39. It is unlikely that Jane will make critical errors in her job in the future.
40. Jane is a valuable member of the department.

**Please give your ratings on a 9 point scale with 9 indicating the highest positive rating and 1 indicating the lowest negative rating.**

41. How would you rate Jane's overall performance.
42. How would you rate Jane's interpersonal skills.

43. How would you rate Jane's potential for future promotion.
  44. How would you rate Jane's value as a team player in the department.
  45. How would you rate Jane's work habits.
  46. How would you rate Jane's innate abilities and aptitudes.
  47. How would you rate Jane's level of work motivation.
  48. How would you rate Jane's effort level.
  49. How would you rate the contribution of Jane's personality to work place effectiveness.
  50. What is your overall evaluation of Jane as a worker.
  51. **As the Ombudsperson you are asked to select one of the following actions. The company has to abide by your decision.**
    1. Reinstatement Jane fully and make up for lost pay and benefits.
    2. Reinstatement Jane fully and make up for lost pay and benefits but give her an oral warning.
    3. Reinstatement Jane fully and make up for lost pay and benefits but put a written warning in her personnel file.
    4. Reinstatement Jane, and reduce the discharge to a suspension of two days without pay, and put a written warning in her personnel file.
    5. Reinstatement Jane, and reduce the discharge to a suspension of one week without pay, and put a written warning in her personnel file.
    6. Reinstatement Jane, and reduce the discharge to a suspension of three weeks without pay, and put a written warning in her personnel file.
    7. Uphold Jane's termination.
- Briefly, give reasons for your decision.

## THE CASE OF THE GRIEVANT

Joe, 24 years old and single, works in the claims processing department for the National Insurance company at one of its major regional offices. Joe is considered to be a warm and expressive individual with a congenial personality. He has a two year technical degree in data processing, and joined the company about a year and a half ago after receiving his degree. His job is to enter vital data about claims that are being processed or that have been settled into the computer. This includes statements made by all the involved parties, the claim adjuster's assessment of the situation, the potential liability associated with the claim, and sometimes the amount of money requested to obtain a release. There are approximately 15 other claims processors who work in the department. Two senior claims processors (Betty and Sam) assist the others, and train the new employees on various facets of the job if necessary. The department is run by the claims processing manager, who allocates the work and gives specific instructions.

The atmosphere in the department in the past generally has been informal and friendly. Disputes between supervisors and workers are pretty rare. The company has provisions for a formal appeal, if a worker is not satisfied with the decision of a supervisor. This policy is national in its scope and covers all offices and departments in various regions of the country. The appeals procedure involves the selection of a company ombudsperson who specializes in resolving grievances. The company ombudsperson is typically a manager who is not familiar with the grievant or his supervisor, and is completely neutral in the dispute. The decision of the ombudsperson is final as far as the company is concerned, and both the grievant and the supervisor must abide by it. National Insurance company is very proud of its ombudsperson program. It holds yearly training sessions that last for 2 weeks to train managers who want to serve in this capacity. The ombudspersons typically have around 15 years of seniority with the company along with reputations for fairness and impartiality in rendering decisions. Employees consider the management of National Insurance company to be very progressive, and feel they are treated fairly. Joe has been very happy with his experience at the National Insurance company until now. A series of incidents over the last 6 months have landed him into trouble. The facts of his situation follow.

Six months ago, upon the retirement of the previous manager Henry, a new manager of the claims processing department was hired. Stanley, the new manager is 45, has a B.A in finance, and has been working in the insurance industry for 23 years. He has had experience working as an insurance salesman as well as a claims adjuster.

The first thing Stanley did after being hired as a manager was to review prior performance of the people in the department using the previous manager's written general assessment of the staff. He held individual meetings with everyone in the department and



told them of his expectations as well as the goals they should shoot for. This created some initial stress in the department as the previous manager Henry only communicated about performance with the staff when there were serious problems, or during the routine performance review. Under Henry, performance review was usually a once a year, 15 minute informal conversation, going from topic to topic. Henry was always easy going and relaxed, and usually ended all his performance review sessions with comments like, "keep up the good work," "you are doing fine" etc.

Stanley, however, has a completely different style of management and the claims processing staff is attempting to adjust to it. Joe in particular became very concerned after his first meeting with Stanley. He doesn't know what others were told, but Stanley told him that his performance as judged by the previous manager Henry was marginal. Stanley said that according to the written general comments made by Henry to bring him up to date on the staff, Joe had a tendency to be easily distracted socially at the work place, and often had to be instructed several times before he adequately grasped a new task. Other information Stanley had obtained on productivity indicated that Joe's output, the number of claims processed per week, was in the bottom 10% of the staff. Stanley told Joe that his performance would be closely monitored, and that he should be making an effort to improve his performance.

Joe was surprised but responded politely to Stanley's comments. He told Stanley that this was the first time he had heard this, and that Henry had never formally reviewed him. Joe further stated that Henry had only made positive comments to him about his work. He also asked Stanley whether he had taken into account the fact that a lot of the claims he handles tend to be very complex, involving serious accidents, and they take longer to process. Stanley said only that he was working with the raw data and the meeting was to identify general patterns and trends in the department. The meeting ended somewhat awkwardly, with Joe respectfully emphasizing that not only numbers of claims but also the complexity of claims should be looked at for performance review purposes.

Two months later, Stanley called Joe in the office and talked with him again. "I have been told that you are taking 20-30 extra minutes on your lunch hour for the last week. What is going on?" "Nothing", replied Joe. "I have some friends who are visiting, and I have been eating with them. I make sure Betty or Sam (the senior claim processors) know that I am going to be late. I make up the time in the evening." "Well, you really need to focus on your work more. You mixed up information on two claims last week. I just got a call from the manager in the claims adjusting department, who told me that because of the mixup he has to rewrite both releases, which will delay the signing by at least a day. Sometimes, if you don't get those things signed quickly, people change their minds," said Stanley. "Stanley, I processed those claims exactly as I was given by Betty. Paperwork for them probably got mixed up before they even got to me," responded Joe.

"That may very well be, but that is the kind of thing we should be able to catch here. Don't you think so?" said Stanley as he pressed on. "Honestly, I think that's a little unrealistic. I know other people here who have made those kind of mistakes before. They do happen. I am sorry but I am not perfect. I do my fair share of work around here and I do a good job. I don't know what else I can say." said Joe. The meeting ended shortly thereafter. It was clear that relations between Stanley and Joe had become tense.

Three months later, another incident took place that has led to the current dilemma faced by Joe. All the claim processors, typically before leaving for the day, back up the data they have entered to ensure safety. This practice, although required by company policy is not always followed. In fact some claims processors have been known to back up data only once a week on Friday. Last Friday afternoon after lunch, Joe was asked by Sam, one of the senior claim processors, to assist in training a new data-entry person. The training took the whole afternoon, and since Joe was not at his terminal, he forgot to backup the data. He remembered when he got home and gave Sam a call at home to tell him what had happened. "Ah, hell don't worry about it. Do it on Monday morning." Joe went to work as usual on Monday, and started processing new claims. After a couple of hours he went to get some coffee, and when he came back to the terminal, the screen was stationary. Joe hit the keyboard and nothing happened. He realized that something was wrong and called Sam. Sam played with the computer and said that the hard disk had crashed and phoned the technician.

It soon became evident that Joe had lost some very important data, as he had not backed up his work on Friday or his work on Monday morning. Sam assured him it would be O.K. They would simply get the original paper work, and Joe could enter it all in again. When Stanley heard of this, however, he was furious and called a department meeting. At the meeting he said, "I think you all know how important claims data is to the functioning of this company. The people we insure trust us to process their claims and settle them as speedily as possible. In order to do that we have to take great care in collecting and maintaining this data. Frankly, this department, or I should say some people in this department, have developed bad habits regarding how claims data is to be processed and treated. This is a serious job, and we can not be frivolous about it. I want you all to read the company manual on claims processing, especially the section on the importance of backing up data. According to the company policy, an employee responsible for losing claims data can be summarily dismissed."

After the meeting ended, Stanley called Joe in his office. He said, "I am sorry Joe, but I really don't think you belong here. The kind of work we do really requires an aptitude for and mastery of detail, and frankly you just don't have it. I think it will be in your interest to find something more suitable for yourself. I will be happy to give you a recommendation letter." Joe was completely stunned and speechless. After a few

seconds he finally said, Stanley "I don't know what you have against me but you are being completely unfair." Stanley, however, continued, "I have been thinking about this off and on. Your performance is poor in general in this department. I think you are going to be better off, and we are going to be better off." Joe looked at Stanley and said, "I don't know why you are doing this. I honestly don't think I deserve this. This doesn't make any sense. Please think about this. This is a very serious thing for me." Stanley did not say anything. Joe then got up and left Stanley's office. He was quite shaken up and his face was pale. He saw Sam and Betty talking outside. He went up to them and said "Stanley just fired me." "Are you serious," said Sam. "Yes, yes.. I don't know what's going on," said Joe.

Sam and Betty, the two senior claim processors both went in to talk to Stanley. Stanley, however, would not budge. "Joe is not as bad as you think Stan. He gets his work done on time most of the time. I know he is a little social at work sometimes, but he pitches in. He is really helpful around here and will even help train newcomers when I ask him to. That helps us all out. That is what I was having him do when he forgot to backup. Stanley said that he had the discretion to fire workers when he felt their performance was below par. He told Sam and Betty that after observing Joe for 6 months, he felt he was making an informed decision. "You know he can appeal this," said Betty. "You are of course familiar with the appeals procedure at the company," asked Sam. "Look, both of you, I have made my decision. I run this department and I have made my decision. He can appeal it to whomever he wants but this is his last week in this department. I understand your sympathy for him, but my goal is to have a smooth functioning and an efficient department. This means that some people are going to have to start performing better than they have been. If they are unwilling to or can't, then they should find a more suitable position elsewhere."

Joe did appeal the decision. A senior manager, acting as the ombudsperson from a different location was selected to hear the grievance. Put yourself in the position of that manager and indicate the extent of your agreement or disagreement with the following statements on a 9 point scale. **Nine indicates Strongly Agree** and **1 indicates Strongly Disagree**. Please remember that there are no right or wrong answers.

1. Stanley is a fair manager.
2. Joe has the ability to be a good performer.
3. Joe did not back up the data when he was supposed to.
4. Joe's mistake was a freak accident.

5. Joe is likely to make serious mistakes at the workplace in the future.
6. Joe's job is an easy one.
7. Joe gets along well with his co-workers.
8. Joe does not have an aptitude for detail work.
9. Joe contributes significantly to the overall department effectiveness.
10. The cases assigned to Joe are complex.
11. Joe's future performance is likely to be good.
12. Joe is careless at work.
13. Joe successfully accomplishes the tasks assigned to him.
14. Joe is not required to backup the data as part of his job.
15. Joe is a conscientious worker.
16. Joe does not come up to the standards of the department.
17. Joe has the responsibility for saving data at the proper time.
18. Joe has no one to blame but himself for his predicament.
19. Joe is easily distracted from his work.
20. The work performed in Joe's department requires specialized training.
21. Stanley's decision regarding Joe lacks basic fairness.
22. Joe's failure to backup the data was due to a factor outside his control.
23. Joe is competent in what he does.
24. Joe has good work habits.
25. Stanley's decision to fire Joe is an overreaction.

26. The type of work Joe performs is relatively routine.
27. Joe does not put enough effort in his work.
28. What happened to Joe could have happened to anybody.
29. Stanley made a well thought out decision when he fired Joe.
30. Joe is a highly skilled worker.
31. Bad Luck played a major role in Joe's predicament.
32. Joe works hard at his job.
33. Joe is helpful at the work place.
34. Joe has a demanding job.
35. Joe is a good performer at the work place.
36. Joe puts a lot of energy in his work.
37. Joe's mistake reveals something about Joe that is unlikely to change in the future.
38. Joe's present situation is a result of Stanley's unfairness.
39. It is unlikely that Joe will make critical errors in his job in the future.
40. Joe is a valuable member of the department.

**Please give your ratings on a 9 point scale with 9 indicating the highest positive rating and 1 indicating the lowest negative rating.**

41. How would you rate Joe's overall performance.
42. How would you rate Joe's interpersonal skills.
43. How would you rate Joe's potential for future promotion.
44. How would you rate Joe's value as a team player in the department.

45. How would you rate Joe's work habits.
46. How would you rate Joe's innate abilities and aptitudes.
47. How would you rate Joe's level of work motivation.
48. How would you rate Joe's effort level.
49. How would you rate the contribution of Joe's personality to work place effectiveness.
50. What is your overall evaluation of Joe as a worker.
51. **As the Ombudsperson you are asked to select one of the following actions. The company has to abide by your decision.**
  1. Reinstatement Joe fully and make up for lost pay and benefits.
  2. Reinstatement Joe fully and make up for lost pay and benefits but give him an oral warning.
  3. Reinstatement Joe fully and make up for lost pay and benefits but put a written warning in his personnel file.
  4. Reinstatement Joe, and reduce the discharge to a suspension of two days without pay, and put a written warning in his personnel file.
  5. Reinstatement Joe, and reduce the discharge to a suspension of one week without pay, and put a written warning in his personnel file.
  6. Reinstatement Joe, and reduce the discharge to a suspension of three weeks without pay, and put a written warning in his personnel file.
  7. Uphold Joe's termination.

Briefly, give reasons for your decision.

**FOR ITEMS 1-40 PLEASE USE THE FOLLOWING SCALE.**

1-----2-----3-----4-----5-----6-----7-----8-----9  
**STRONGLY STRONGLY**  
**DISAGREE AGREE**

**FOR ITEMS 41-50 PLEASE USE THE FOLLOWING SCALE.**

1-----2-----3-----4-----5-----6-----7-----8-----9  
**LOWEST HIGHEST**  
**NEGATIVE POSITIVE**  
**RATING RATING**

**FOR ITEM NO. 51 PLEASE USE THE FOLLOWING SCALE TO MAKE THE DECISION. YOU NEED TO PICK ONE OPTION FROM 1 TO 7. AFTER YOU PICK THE OPTION, MARK IT ON YOUR SHEET.**

1-----2-----3-----4-----5-----6-----7

**THE FOLLOWING DEMOGRAPHIC DATA IS REQUESTED. PLEASE CIRCLE THE APPROPRIATE ANSWER OR WRITE IN IF NEEDED.**

**1. YOUR GENDER: FEMALE MALE**

**2. YOUR RACE: BLACK WHITE OTHER**

**3. THE DISTRICT THAT YOU ARE IN (PLEASE WRITE THIS IN)**

**4. YOUR PROGRAM AREA**

**AGRICULTURE**

**HOME ECONOMICS**

**COMMUNITY RESOURCE DEVELOPMENT**

**4H**

**5. THE NUMBER OF YEARS YOU HAVE BEEN A UNIT DIRECTOR (PLEASE WRITE THIS IN)**

**6. THE HIGHEST DEGREE THAT YOU HAVE (PLEASE WRITE IN).**

**7. YOUR DATE OF BIRTH**

**8. YOUR NAME:**

**THANK YOU FOR YOUR HELP IN DOING THIS. INFORMATION GAINED FROM THIS STUDY WILL BE SHARED WITH YOU ON THE MAY 24 MEETING OF THE UNIT DIRECTORS.**



**HARSH LUTHAR  
DEPARTMENT OF MANAGEMENT  
BRYANT COLLEGE  
SMITHFIELD, RI 02917**

**EDUCATION**

**Ph.D.** - Department of Management, The R.B Pamplin College of Business, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0233.  
**GPA-3.98.** Elected to Membership of The Honor Society of Phi Kappa Phi on April 8, 1991.

**Doctoral Seminars:** Organization Behavior, Human Resources Management, Labor Relations, Research Methods/Statistics.

**Dissertation Topic:** Two-Way Gender Bias in Management Evaluations and Decision Making.

**Dissertation Chair:** Dr. T.W. Bonham, Professor of Management at the R.B. Pamplin College of Business, VA. TECH., Blacksburg, VA.

**MBA** - University of Wisconsin-Whitewater, December, 1984.

**B.A** - Beloit College, Beloit, Wisconsin, May, 1978.

**EDUCATIONAL WORK EXPERIENCE**

**Assistant Professor** - Department of Management, Bryant College, August 1993 to Present. Teaching Human Resources and Organization Behavior.

**Instructor** - Department of Management, Virginia Tech, Summer 1990, Spring 1992.  
Labor Relations.

**Research Assistant** - Department of Management, Virginia Tech, 1989 - 1991.

**Instructor** - Longwood College, Farmville, Virginia. August 1986 to May 1989.  
Courses taught: Human Resources Management, Industrial Relations,  
Organization Theory, Production/Operations Management.

**Instructor**, Madison Area Technical College, Madison, WI, 1985-86.

Taught Business Statistics.

**Lecturer**, University of Wisconsin Center, Janesville, WI, 1985-86.

Taught Business Statistics.

### **RESEARCH AND PUBLICATIONS**

Michael J. Dougherty and Harsh K. Luthar. "Wage Negotiations after the Fort Stewart Decision: An Opening for Federal Employees or a Mirage?." Proceedings of the 44th Annual Meeting, Industrial Relations Research Association, 1991. Abstract, In Press.

Harsh K. Luthar and Jerald F. Robinson. "Arbitration of Title VII Disputes in Academe: Impact of University of Pennsylvania v. EEOC." Proceedings of the 43rd Annual Meeting, Industrial Relations Research Association, 1990, pp. 416-424.

Dane M. Partridge, Kent F. Murrmann, Harsh K. Luthar, and Gordon Cheung. "Union Organizing in the South: Progress and Prospects." Proceedings of the Southern Industrial Relations Research Association, 1989. Published Abstract of a Professional Paper Presentation, pp. 70-72.

### **RESEARCH IN PROGRESS**

- A) Investigation of the two-way gender bias in management evaluations and decision making (Dissertation).
- B) Investigating how different measurement procedures used in management research (such as employee attitude surveys) can produce differing results. In particular the Likert scale is being compared to magnitude and continuous scaling to discover the differences in reliability when measuring the same attitude objects. Data analysis stage.

### **NON-ACADEMIC WORK EXPERIENCE**

**Operations Manager**, Headwaters Distributing Company, Eden Prairie, MN. January, 1985 to June, 1985. Responsibilities included managing accounts receivables and payable, inventory planning and control, purchasing of supplies and equipment, and supervision of employees in the warehouse and office.

**Health Club Instructor**, New York Health and Racquet Club. February, 1979 to August 1981. Taught exercise classes and sold health club memberships. Promoted to

Floor Manager after a year.

**Proofreader**, Skadden, Arps, Slate, Meagher and Flom-New York, N.Y. November 1977 to April 1978. Internship through Beloit College.

### **MISCELLANEOUS**

**Founder** and Co-advisor to the Longwood Chapter of Society for the Advancement of Management (S.A.M), 1986-1989.

**Member** of Delta Sigma Pi, a professional business fraternity.

**President**, Graduate Business Association, UW-Whitewater, 1983-84 academic year.

**Broadcaster**, Business Update News at WSUW Radio Station, Spring 1984.

**Leader**, Student Orientation, UW-Whitewater for summer 1982.

### **PERSONAL INFORMATION**

Born 9/12/56 to Raj and Rajendra Luthar

Married to Anshu Luthar on December 28, 1986

Father of Jay Luthar born on March 5, 1988

A handwritten signature in cursive script that reads "Harsh Luthar". The signature is written in black ink and is positioned to the right of the personal information text.