2017

Why Can't a Woman Fail Like a Man? Gender Differences in Perceived Competence Following a Mistake

Kathryn Kincaid
Butler University

BUTLER UNIVERSITY HONORS PROGRAM

Honors Thesis Certification

Please type all information in this section:

Applicant Kathryn Kincaid

(Name as it is to appear on diploma)

Thesis title Why Can't a Woman Fail Like a Man? Gender Differences in Perceived Competence Following a Mistake

Intended date of commencement May 6, 2017

Read, approved, and signed by:

Thesis adviser(s) _____ 3/17/17
Date

_____ Date

Reader(s) _____ 4/26/17
Date

_____ Date

Certified by _____ _____
Director, Honors Program Date

Why Can't a Woman Fail Like a Man? Gender Differences in Perceived Competence Following a Mistake

A Thesis

Presented to the Department of Psychology

College of Liberal Arts and Sciences

and

The Honors Program

of

Butler University

In Partial Fulfillment

of the Requirements for Graduation Honors

Kathryn Kincaid

April 24. 2017

Abstract

Stereotypes are pervasive and can significantly influence the way we perceive and evaluate others. When people occupy roles that are not congruent with stereotypes (such as a stay-at-home dad or a female CEO), past research has suggested that they are evaluated more harshly than those in roles that are stereotype-congruent. The present study examined the role of gender stereotypes by asking participants to read a vignette about a college student and their performance in a class. In these vignettes, the student's major and gender were manipulated such that there were students in gender stereotype-congruent majors (female nursing major, male computer science major) and students in stereotype-incongruent majors (female computer science major, male nursing major). Participants were then asked to evaluate the student's performance, providing rating of competence, status, and likeability. Analyses revealed that there was no significant effect of stereotype-congruence on evaluations, a finding inconsistent with prior work. This discrepancy is discussed in light of differences in participants' familiarity with the role and job of the person being evaluated.

Why Can't a Woman Fail Like a Man? Gender Differences in Perceived Competence Following a Mistake

Stereotypes are pervasive in everyday life, shaping how humans interact with one another and the world. A stereotype can be defined as an over-generalized, fixed belief about a particular group of people (Kanahara, 2006; McCauley, 1980). Broadly speaking, stereotypes can be conceptualized as oversimplified pictures of the world, which help us satisfy our need to see the world as more predictable and understandable than it really is (Lippmann, 1922).These beliefs can be cognitively advantageous, allowing us to apply knowledge from prior experiences to current similar experiences, respond rapidly, and conserve cognitive resources. Stereotypes simplify our social worlds and provide guidance when navigating new experiences (McCauley, 1980, Lippmann, 1992). However, while contributing to speed and efficiency, stereotypes can cause us to overlook individual differences and make unwarranted assumptions about people based on their membership in a particular group.

Some of the most pervasive stereotypes relate to gender, such as the belief that women are more suited for working with people, while men are more competent with machines. Gender-based stereotypes have both wide and well documented effects on the type of judgments people tend to make about others. Gender stereotypes can lead individuals to believe that certain jobs or careers paths are more appropriate for women and others are more appropriate for males. Those stereotypes can influence others' judgments and perceptions of the people in those roles, especially when they are not consistent with a stereotype about the role. Because gender stereotypes are so pervasive and influence our perception of others so powerfully, individuals who hold positions or roles that are inconsistent with these stereotypes are often perceived in negative ways.

People in positions that are inconsistent with stereotypes (counterstereotypic positions) often face discrimination, in large part because stereotypes have immense effects on the type of judgments people tend to make about others (Heilman, Wallen, Fuchs, & Tamkins, 2004; Heilman & Okimoto, 2007; Eagly

& Karau, 2002; Rudman & Glick, 1999). Past research emphasizes dislike as a potential mechanism for this discrimination, in that individuals in counterstereotypic positions are often personally disliked. For example, women who are successful in traditionally male tasks are perceived to be violating gender-stereotypic prescriptions and are penalized for this violation. These women are perceived as less competent than their male peers and are often disliked or interpersonally derogated (e.g., Heilman, Wallen, Fuchs, & Tamkins, 2004; Heilman & Okimoto, 2007). Similarly, leadership roles are largely held by men, are associated with the male gender role, and often require stereotypically male or agentic behavior. As a result, women in leadership roles tend to be perceived as cold, unlikeable, and interpersonally deficient (e.g., Eagly & Karau, 2002; Rudman & Glick,1999). These negative evaluations arise not from generalized negative attitudes women, but perceived incongruence between gender stereotypes and women's behavior (Eagly & Karau, 2002).

Women are often described as facing a "glass ceiling," wherein they are less likely than men to reach leadership positions. In addition to facing a glass ceiling, women may also be positioned on a "glass cliff," as women who do work their way up to leadership positions are more likely than men to subsequently lose their these positions (Ryan & Haslam, 2005). One explanation for this effect is that women appear more likely to achieve leadership roles in particularly tumultuous times when the chances of success are low. Salient examples of this phenomenon include Theresa May's ascension to British prime minister following the Brexit vote, Carly Fiorina's brief tenure with HP followed by an unsuccessful presidential candidacy in a tumultuous race, or the tendency for Fortune 500 companies to be most likely to appoint female or minority CEOs when their companies are facing a downturn. While being appointed to precarious positions presents one explanation, another potential mechanism to explain glass cliff effects is that the fragility of counterstereotypic individuals' reputations.

For individuals in gender stereotype-incongruent positions, making small mistakes on the job is especially damaging. Brescoll, Dawson, & Uhlmann (2010) examined how people make judgements of

men and women in both gender-congruent and gender-incongruent leadership jobs. When men and women in gender-incongruent positions perform their jobs without mistake, they were perceived as equally competent. However, when individuals with gender-incongruent positions made a mistake on the job they were perceived as significantly less competent than were their peers in gender-congruent positions. This suggests that although individuals in gender-incongruent positions can be successful and achieve high status, this status is fragile and unstable. Individuals in counterstereotypic positions may experience sharper penalties for their mistakes than their stereotype-congruent peers.

The status of counterstereotypic individuals is likely fragile because stereotypes are especially influential in instances of ambiguity. For example, minority job candidates with unambiguously strong qualifications are generally evaluated fairly, but those with mixed or ambiguous track records are often discriminated against when compared with white candidates (Hodson, Dovidio, & Gaertner, 2002). When leaders make mistakes on the job, this creates ambiguity as to their competence and provides an opening for stereotypes to influence perceptions. Thus, when individuals in counterstereotypic leadership positions make a mistake and create ambiguity, they will be evaluated more negatively than those in stereotype-congruent positions.

Prior research has focused on perceptions of men and women in gender-incongruent leaderships positions, often using positions with which undergraduate participants have little direct experience. For example, Brescoll et al. showed that counterstereotypic individuals are judged more harshly when they made a mistake, but the targets to be evaluated held the role of police chief or women's college president. These are roles that most undergraduates have little contact with, and generally lack knowledge about these leaders' duties and what might constitute a reasonable mistake on the job. In cases where undergraduates have little direct experience with the target role they are evaluating, it might be expected that undergraduate participants would rely more heavily on stereotypes when evaluating the leaders.

To extend prior work, the current study sought to examine whether individuals in counter-stereotypic positions are evaluated more negatively when the evaluator is familiar with the position. Using a role undergraduates are familiar with, the current study examined perceptions of students in both gender congruent and gender-incongruent majors. This allowed participants to evaluate presumed peers, using tasks and mistakes with which undergraduates are more familiar. Further, the current study attempted to extend the past literature by examining perceptions of individuals' competence and likeability in gender-congruent versus incongruent positions not only when they fail, but also when they excel or perform exceptionally well in their field.

The purpose of this study was to examine how college students' successes and failures are differentially perceived based on whether they are in a gender stereotype-congruent major (e.g., a female nursing major) or a gender stereotype-incongruent major (e.g., a female computer science major). To achieve this goal, participants read vignettes about a college student's performance in a class, where the student's gender, major, and level of success were manipulated. Participants then evaluated this target student using measures of competence, status, and likeability. Based on prior research, I hypothesized that:

1) When students in a stereotype-incongruent major make a mistake, they will be perceived as less competent, less likeable, and will be conferred less status than students in a stereotype-congruent major who make the same mistake.
2) Students who are successful in a stereotype-incongruent major will not be perceived as competent, likeable, or conferred the same level of status as students who are successful in a stereotype-congruent major.
3) When students do not excel or make a mistake (their performance is simply adequate), students in stereotype-congruent and stereotype-incongruent majors will be perceived as equally competent, equally likeable, and will be conferred equivalent status.

Method

Pilot Study

All participants read a brief vignette about a college student and their performance in class. The specific majors and performance conditions used in the vignettes were pilot tested to ensure that the chosen majors were sufficiently gender-stereotyped and that the example mistakes for each major were seen as equally severe, ensuring that the mistake presented for the female- or male-typed major did not appear more or less severe than the mistake for the opposite major. For the pilot study, 40 Butler University undergraduates completed an online survey and received extra credit in a psychology course in exchange for their time. The majority of the participants were psychology majors or students of other majors who were enrolled in an introductory psychology course.

Each participant in the pilot study read three different scenarios about a college student. These consisted of one scenario where the student failed, one where the student performed adequately, and a final scenario where the student excelled. Academic major was manipulated as a between-subjects variable such that different participants read scenarios about students in different majors. All language in the vignette was gender-neutral, ensuring that the target student was not identified as male or female. After reading the scenarios, participants were asked to rate the performance of the student on a 5-point Likert scale from very unsuccessful to very successful. They were then asked to rate the prestige of four different college majors that had been identified for possible use in the study (nursing, education, computer science, and sports communication). After those ratings, participants estimated based on their experience the gender makeup of students in each major using a provided 5- point Likert scale that ranged from 1= 10% men and 90% women to the reverse 5= 10% women and 90% men.

Based on the results of the pilot study, computer science and nursing were selected as the two majors for inclusion in the main study. The average estimated gender ratios were equally skewed for the selected female-typed major (nursing, estimated to be 75% female) and male-typed major (computer

science, estimated to be 75% male), (t(38) = 0.84, n.s.). There was no significant difference in the perceived prestige of nursing (M= 3.79) and computer science (M = 3.85), (t(38) = .388, n.s.). Finally, success and failure were rated equivalently across these two majors. There was no difference in the perceived success of the computer science major (M = 3.89) and the nursing major (M = 4.14), (t(38) = 1.42, n.s.), or the perceived failure of the computer science major (M = 1.56) and the nursing major (M = 1.96), (t(38) = 0.47, n.s.).

The names for the target students (Jake and Emily) used in the actual study to manipulate gender were chosen by identifying the birth cohort of likely participants (1995) and locating the most popular male and female baby names given that year. This was done to ensure that the names were high frequency and unambiguously referred to someone who was male (Jake) or female (Emily).

Participants

Participants in this study were 330 people who completed the research study online. The majority were Butler University undergraduate students who received extra credit in a psychology course in exchange for their time. Participant demographic information was not collected, but the students were sampled from a university population that is largely traditional college age (18-22), majority female, and predominately white. The remainder of the participants were recruited online through the Social Psychology Network and Psychological Research on the Net.

Materials

All participants read a brief (62-word) vignette about a college student and their performance in a class. The student either excelled (received the highest grade in their class on an exam), performed adequately (received a passing grade on a midterm) or failed (received a failing grade on the midterm). These student's gender and major were manipulated such that there were students with gender-incongruent majors (e.g., a female computer science major or male nursing major) as well as

students with gender-congruent majors (e.g., female nursing major or male computer science major). The following is the scenario read by participants in the female/computer science/failure condition.

> Emily is a senior computer science major who is currently enrolled in the last computer science class required to graduate. For this class, the midterm exam score will make up 30% of her final course grade. However, Emily fails to adequately prepare for the exam and receives a failing grade. Emily is no longer passing the class, which is required to graduate.

The 11 other scenarios used in the study had equivalent text with manipulations of the name (Jake for Emily), the major (Nursing for Computer Science) and the performance. For the neutral performance condition, the last sentence read "Emily (Jake) prepares adequately for this exam and receives an average grade. Emily (Jake) is passing the class, which is required to graduate." For the success condition, this last sentence read "Emily (Jake) prepares extensively for the exam and does very well, earning the highest grade in the class."

After reading the assigned scenario, participants were asked to evaluate the target student on a series of traits. Specifically, competence, status, and warmth served as the three primary dependent measures. Each measure was assessed with a scale composed of two seven-point Likert-type items. The perceived competence scale ($\alpha = 0.82$) consisted two items: a rating of the student's knowledge (very ignorant to very knowledgeable) and competence (very incompetent to very competent). On the scale of likeability ($\alpha = 0.72$), participants were asked to rate how warm they felt towards the student (very cold to very warm) and how likeable the student was (very not likeable to very likeable). On the final scale of perceived status ($\alpha = 0.81$), participants were asked to rate how likely the student was to have a successful career in their chosen field (very unlikely to very likely) and the extent to which they would like to collaborate with the student on a group project.

Procedure

Participants completed the study using a web-based interface through LimeSurvey. Participants recruited through Sona Systems, Social Psychology Network, and Psychological Research on the Net were directed to a hyperlink where they first read an informed consent document. Once they indicated consent, participants were randomly assigned to one of 12 conditions. These conditions corresponded to the 12 unique scenarios, which were generated by manipulating the three between-subjects independent variables of scenario gender (male or female), scenario major (computer science or nursing), and scenario performance (failure, neutral, or success). Participants were assigned to a condition using a random number generator which produced a number between one and 12 which in turn corresponded to a specific condition. The web-based presentation system was designed so that the presented vignette was dependent on this random number generator, and participants were only shown the one that corresponded with their assigned condition. Participants were asked to read the vignette that was presented to them and then to respond to six questions about how they viewed the student in the vignette. Following completion of the rating scales, participants were thanked for their time and, for the students participating from Butler University, directed to a link to claim extra credit. The entire procedure took less than 15 minutes.

Design

This study was a 2 (Scenario major: nursing or computer science) X 2 (Scenario gender: male or female) X 3 (Scenario performance: failure, neutral, or success) design, with the three primary dependent measures being perceived competence, status, and likeability.

Results

A series of 2 (Scenario major) X 2 (Scenario gender) X 3 (Scenario performance) ANOVAs were conducted, one for each of the dependent variables: competence, likeability, and status. Means and standard deviations for each measure broken down by each factor are presented in Tables 1 through 3 respectively.

As can be seen in Table 1 for the dependent variable of Perceived Competence, there was a significant main effect of performance, such that students who had succeeded (earned the highest grade in their class on a midterm exam) were viewed as significantly more competent (M = 6.48, s = 0.85) than students who had performed adequately (passed the midterm exam) (M = 5.58, s = 0.96) and students who had failed (failed the midterm exam) (M = 3.98, s = 1.28), ($F(1, 285) = 127.10$, $p < 0.001$, $\eta p^2 = 0.482$). Post hoc analyses using the Scheffé post hoc criterion for significance indicated that there were significant differences in perceived competence between all three performance conditions. There was no main effect for gender, ($F(1, 285) = 1.094$, n.s,), or major ($F(1, 285) = 1.09$, n.s.). Thus, the target student's gender and major did not significantly affect perceived competence. There were no significant interactions between these variables.

As can be seen in Table 2 for conferred status, there was again a significant main effect of performance ($F(1, 286) = 106.59$, $p < 0.0001$, $\eta p^2 = 0.470$) such that students who succeeded were conferred significantly greater status (M = 6.10, s = 1.15) than students who performed adequately (M = 5.08, s = 1.35) or failed (M = 3.35, s = 1.17). Post hoc analyses using the Scheffé post hoc criterion for significance again revealed that there were significant differences in Perceived Status between all three performance conditions. There was no main effect for gender or major, and no significant interaction between these variables.

As can be seen in Table 3 for Likeability, there was a significant main effect of performance, such that the target students who succeeded were perceived as significantly more likeable (M = 5.38, s = 0.98) than students who performed adequately (M = 5.12, s = 0.99) or failed (M = 4.23, s = 0.97) ($F(1, 285) = 32.65$, $p < 0.0001$, $\eta p^2 = 0.182$). Again, post hoc analyses using the Scheffé post hoc criterion for significance revealed that there were significant differences in likeability between all three performance conditions. For the dependent variable of likeability, there was also a significant main effect of gender ($F(1, 285) = 6.51$, $p < 0.01$, $\eta p^2 = 0.018$) such that female students were rated as significantly more

likeable (M = 5.03, s = 1.15) than male students (M = 4.79, s = 1.02). There was no main effect of major, and no significant interactions.

Discussion

The three primary hypotheses tested in this study were that 1) when students in a stereotype-incongruent major make a mistake, they would be perceived as less competent, less likeable, and conferred less status than students in a stereotype-congruent major who make the same mistake, 2) when students in a stereotype-incongruent major are successful, they would not be perceived as competent, likeable, or conferred the same level of status as students who are successful in a stereotype-congruent major, and 3) when students do not excel or make a mistake (their performance is simply adequate), students in stereotype-congruent and stereotype-incongruent majors will be perceived as equally competent, equally likeable, and will be conferred equivalent status. None of these hypotheses were supported, as the stereotype-congruence of a student's' major had no effect on perceived competence, likeability, or status.

A manipulation check did confirm that the manipulation of vignette student performance was successful, as performance had a significant effect on all evaluations, regardless of the target student's major or gender. Participants did perceive that the student in the success condition had succeeded, and that the student in the failure condition had failed. Further, a major strength of the present study was the large sample size (N= 333). Given this large sample, it is unlikely that a real difference between perceptions of gender differences for individuals in stereotype congruent and incongruent roles exists but went undetected. Finally, there were no trends in the nonsignificant results, that is to say that no group differences nor the hypothesized interaction even approached significance. Therefore, we can be confident in the observed nonsignificant results and interpret them as meaningful in their own right.

These results are inconsistent with prior work, which has shown that the stereotype-congruence of a target's role has significant effects on how a target is evaluated following a failure (Brescoll et al, 2010).

However, this prior work had relied on undergraduate participants evaluating targets in roles with which they are unfamiliar, and with which they have likely had little to no direct contact. In the current study, participants evaluated a target in a familiar role. Undergraduate student participants evaluated their peers, allowing the roles, successes, and failures to all be familiar. The successes and failures (failing an exam and earning the highest score on an exam) are situations that undergraduate participants have likely experienced themselves, or have direct contact with someone who has experienced these successes or failures. While the participants were not entirely undergraduates, the remainder were recruited from online studies that presumably draw from a college student and college-educated population, so again, these participants would also be familiar with the target role.

The discrepancy between prior work that showed a significant influence of stereotype-congruence in evaluation and the current study which showed no influence can be explained by differences in familiarity. Stereotypes are a means of making our world more simple and understandable than it really is (Lippmann, 1922), so they are relied on more heavily in times when the world is not understandable. When people are asked to evaluate someone in an unfamiliar role, the world is not particularly understandable and people rely on stereotypes to form their opinions. However, when people are evaluating someone in a familiar role, they have more knowledge and experience to draw upon, and avoid using stereotypes to formulate evaluations.

While past research has emphasized the ways in which stereotypes cloud our evaluations of others and are especially detrimental to people who do not fit with traditional stereotypes, the present study highlights an area of equality: when people are familiar with the roles and jobs of others, they are more likely to be fair and unbiased in their evaluations. These results underscore the importance of familiarity, contact, and connecting with people who take on roles different than our own.

Limitations of the current study include the lack of demographic variables, as an analysis of evaluations by participant gender might have been a useful addition to aid in explaining the observed

results. Further, the evaluations were formulated on the basis of a written vignette, a procedure that is relatively low in ecological validity. In real-life settings, people more often evaluated one another on the basis of face-to-face interactions, or through formalized written documents such as resumes and CVs.

In future work, I aim to address these limitations and examine the explanation of differences in familiarity by replicating work that showed that people rely on stereotypes to evaluate a target in an unfamiliar role. Using the same methodology as used in the current study, this work will use vignettes of unfamiliar roles (a police chief and women's college president), as was used in prior research. If this future work replicates past findings, then there will exist solid evidence that the stereotype-congruence of a target's role significantly affects evaluations of that target only when the role is unfamiliar; however, stereotype-incongruence has no effect on evaluations when the role is familiar.

References

Brescoll, V. L., Dawson, E., & Uhlmann, E. L. (2010). Hard won and easily lost: The fragile status of leaders in gender-stereotype-incongruent occupations. Psychological Science, 21(11), 1640-1642.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. Psychological review, 109(3), 573.

Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks?: The implied communality deficit. Journal Of Applied Psychology, 92(1), 81-92.

Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for Success: Reactions to Women Who Succeed at Male Gender-Typed Tasks. Journal Of Applied Psychology, 89(3), 416-427

Hodson, G., Dovidio, J. F., & Gaertner, S. L. (2002). Processes in racial discrimination: Differential weighting of conflicting information. Personality And Social Psychology Bulletin, 28(4), 460-471.

Kanahara, S. (2006). A review of the definitions of stereotype and a proposal for a progressional model. Individual Differences Research, 4(5), 306-321

Lippmann, W. (1922). *Public Opinion.* New York: Harcourt, Brace and Co.

McCauley, C., Stitt, C. L., & Segal, M. (1980). Stereotyping: From prejudice to prediction. Psychological Bulletin, 87(1), 195-208. doi:10.1037/0033-2909.87.1.195

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle managers. Journal Of Personality And Social Psychology, 77(5)

Table 1
Means (and Standard Deviations) of Perceived Competence broken down by Scenario Major, Scenario Gender, and Scenario Performance.

	Computer Science		Nursing	
	Female	Male	Female	Male
Failure	4.077 (1.137)	3.827 (1.363)	4.000 (1.414)	4.026 (1.274)
Neutral	5.865 (0.805)	5.647 (0.580)	5.625 (0.827)	5.275 (1.193)
Success	6.735 (0.472)	6.400 (0.935)	6.250 (1.240)	6.565 (0.550)

Table 2
Means (and Standard Deviations) of Likability broken down by Scenario Major, Scenario Gender, and Scenario Performance.

	Computer Science		Nursing	
	Female	Male	Female	Male
Failure	4.346 (1.056)	4.000 (0.990)	4.476 (1.100)	4.105 (0.591)
Neutral	5.365 (1.045)	4.677 (0.865)	4.938 (0.873)	5.150 (0.988)
Success	5.500 (1.212)	5.180 (0.840)	5.639 (1.054)	5.304 (0.901)

Table 3
Means (and Standard Deviations) of Perceived Status broken down by Scenario Major, Scenario Gender, and Scenario Performance.

	Computer Science		Nursing	
	Female	Male	Female	Male
Failure	3.327 (1.058)	3.365 (1.110)	3.524 (1.487)	3.184 (1.070)
Neutral	5.324 (1.270)	5.235 (1.002)	4.647 (1.539)	4.975 (1.463)
Success	6.147 (0.766)	6.120 (0.950)	6.000 (1.698)	6.130 (1.120)