

The Romance of Leadership: Rekindling the Fire through Replication of Meindl and Ehrlich

Michelle M. Hammond*
Oakland University
School of Business Administration
410 Elliott Hall
275 Varner Drive
Rochester, MI
48309
USA

michellehammond@oakland.edu

*Corresponding Author

Birgit Schyns
Neoma Business School, Campus Reims
Center for Leadership and Effective Organizations
59 Rue Pierre Taittinger
51100 Reims
France

birgit.schyns@neoma-bs.fr

Gretchen Vogelgesang Lester
Lucas College and Graduate School of Business
School of Management
San Jose State University
One Washington Square
San Jose, CA
95192
USA

gretchen.lester@sjsu.edu

Rachel Clapp Smith
Purdue University Northwest
College of Business
2200 169th Street
Hammond, IN
46323
USA

rachel.smith@pnw.edu

Jane Shumski Thomas
Purdue University Northwest
College of Business
2200 169th Street
Hammond, IN
46323
USA
jane.thomas@pnw.edu

Abstract

Given changes in business and society, the romance of leadership theory, which describes a glorification of the perceived influence of leaders on organizational outcomes, is arguably more relevant than at its conception over thirty years ago. This paper presents four studies aimed to replicate Meindl and Ehrlich's (1987) early experiment on the romance of leadership, specifically considering the effect of leadership attributions on company evaluations. Studies 1 and 2 are close replications, whereas Studies 3 and 4 provide a conceptual replication drawing from a broader sample in age and work experience and include additional experimental conditions. These conditions vary the gender of the leader and including both success and failure situations, as well as including additional outcomes variables of participants' behavioral intentions to support, invest, seek employment, or purchase from the company. Taken together, these studies do not support Meindl and Ehrlich's findings that organizations are viewed more favorably when such outcomes are attributed to leadership. We discuss implications for the romance of leadership theory.

Key Words

Romance of leadership; gender; attribution

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**THE ROMANCE OF LEADERSHIP: REKINDLING THE FIRE THROUGH
REPLICATION OF MEINDL AND EHRLICH**

Over thirty years ago, Meindl, Ehrlich, and other colleagues (1985; 1987) proposed an “unconventional, even radical approach that challenged us to critically examine the prevailing evidence for the importance of leadership” (Bligh, Kohles, & Pillai, 2011, p. 1058). The Romance of Leadership (ROL) theory challenged assumptions about how much leadership matters and the extent to which the impact of leadership is overstated. Specifically, the theory highlights an over-attribution of the responsibility for organizational outcomes to leaders vis-à-vis employees, industry, or external forces. Their findings indicated an outsized leader influence—that is, leadership and leaders are romanticized as being the primary drivers of organizational outcomes. ROL is an attributional bias towards the importance of leadership in the functioning of organizations and the ability of leaders to control and influence organizational outcomes (Meindl & Ehrlich, 1987). In essence, ROL is a cognitive shortcut that offers a simple explanation (leaders) for the complex process of organizational performance, with vast implications for both the theory and practice of leadership.

Despite a relatively small body of empirical research, especially in comparison to many other theories of leadership (transformational leadership, servant leadership, etc.), ROL theory “has been highly influential and widely cited” (Collinson, Jones, & Grint, 2018, p. 1627). A search of the phrase “Romance of Leadership” entered into Google Scholar yields over 4,000 results in mid-2021. ROL theory called for less emphasis on identifying effective leader characteristics and behaviors and more on leadership as a sense-making process used to determine causes of success and failure (Bligh et al., 2011; Meindl & Ehrlich, 1987). In doing so, Meindl and colleagues’ work on ROL was arguably a catalyst for the development of post-heroic

leadership theories, such as shared and distributed leadership and the followership theories movement (Fletcher, 2004; Shamir, Pillai, Bligh, & Uhl-Bien, 2007), and challenged charisma, especially, as a hyper-romanticization of leaders (Haslam et al., 2001; Meindl, 1995; Schyns, Felfe, & Blank, 2007; Shamir, 1992). The Meindl and Ehrlich (1987) study was especially pivotal, as it highlighted the impact of the value of leadership as an explanatory power and “this value is high enough to measurably alter the values ascribed to the outcomes it is seen as having produced” (p. 105). In other words, an over-attribution of the influence of leaders on outcomes extends to shape perceptions of the value of the organization more broadly.

According to three of Meindl’s early suggestions, explorations of the questions asked regarding the Romance of Leadership are perhaps even more important today (Bligh et al., 2011), thereby increasing the importance of establishing its replicability. First, Meindl and colleagues (1985) suggested that the tendency for romanticizing leaders is stronger in extreme situations of success or failure. The past 30 years have seen great upheaval; large scale organizational scandals (e.g., the failure of firms such as ENRON, Tyco, and Lehman Brothers; Volkswagen’s Emissions Scandal; Toyota’s Takata airbag recalls), the 2008 global financial crisis and recession, increasing wage disparities, uncertainty related to terrorist activities, and climate catastrophes (Randle, Eckersley, & Miller, 2017), and most recently, the global pandemic of COVID-19 have created instability and stress. As Pillai and Meindl noted in 1998, “crises provide leaders with opportunities to take bold and purposeful action” (p. 647), as the call for leaders is stronger to lead followers out of a crisis. Followers may look to leaders’ optimism, collective support, and directive leadership for hope (Bligh, Kohles, & Meindl, 2004) or they may seek strong autocratic and directive leaders to reduce uncertainty (Harms, Wood, Landay,

Lester, & Lester, 2018; Yukl, 2002). Regardless of the leadership style, followers may have a heroic assumption of leaders in times of uncertainty.

Second, Meindl and colleagues (1985) highlighted the role of the popular press in glorifying leaders. Leaders' narratives are co-constructed with the media and for "high-profile leaders, in particular, the media holds considerable influence over leaders' public image" (Liu, Cutcher & Grant, 2017; p. 696). The rise of social media, with its minimal publishing threshold, has "accelerated the speed at which information is shared, amplified the reach of messages, and solidified the ability of disparate individuals to organize" (Gruber, Smerek, Thomas-Hunt, & James, 2015, p. 164). As the population has access to much more information at a more rapid speed, leaders may be both glorified and vilified through news and social media outlets as it presents a good story.

Finally, Meindl and colleagues (1985) suggest that leaders themselves might contribute to a romanticized view of leadership by creating an "illusion of control." The sophistication of the methods by which leaders craft personae has arguably increased; indeed, much more research into impression management strategies that project idealized images gives potential leaders a blueprint as to how best to reach their constituents (Gray & Densten, 2007; Peck & Hogue, 2018). The frequency and reach of these types of image projections may increase through advances in technology and social media in which leaders can present images and projections instantaneously.

Here, we return to the original assumptions underlying this theory, specifically regarding the impact of attributions individuals make on their evaluations of organizations. We believe it is important to return to these initial conditions, closely replicating the specific attributed impact of leadership on organizational outcomes, and to also conduct a conceptual replication to determine

to what extent additional conditions may influence subsequent evaluations. In addition to the aforementioned impact of ROL and environmental changes potentially increasing the relevance of ROL more generally, several additional reasons motivated our decision to conduct this specific replication study. Although romance of leadership has been examined in different forms, Meindl and Ehrlich (1987) focused on the impact of leadership attribution on company performance. This study has been cited nearly 700 times and many of those papers have been cited in the thousands. These citations provide evidence that this study and the subsequent research into ROL have had a substantial influence on the field of leadership; however, only a few of the assumptions regarding ROL have been explored since the original studies (for exceptions see Felfe & Petersen, 2007; Herrmann & Felfe, 2009; Schyns & Hansbrough, 2012), making a replication even more important. Second, the early claims of ROL theory warrant both close and conceptual replication, given societal changes that may exacerbate the glorification or vilification of leadership. For example, information on company performance and organizational leaders is much more accessible today than it was in 1987. This has likely led to the “Celebrity CEO” phenomena, which can be considered as a form of romance of leadership. Ascribing even minor celebrity status to CEOs can affect investment decisions, stock market performance and overconfidence (Sinha, Inkson & Barker, 2012), we could expect that inflated evaluations of companies through this aspect of romance of leadership could relate to managerial overconfidence, excessive leader compensation, the legitimization of failure, risky investment decisions, and even volatility in the market. With the advent of social media, companies and leaders themselves can contribute much easier to the phenomenon of ROL by highlighting the role of leaders in their organizations (Gray & Densten, 2007).

Overview of Romance of Leadership Theory

The phrase “romance of leadership” was first used in the title of Meindl and colleagues’ (1985) theory-building publication in *Administrative Science Quarterly*, which drew from archival research using public media, academic journals, and undergraduate student interest in leadership and included three experiments. Meindl and Ehrlich (1987) expanded upon these findings with more robust empirical tests of the theoretical model. Following these early studies, the research on the concept of ROL began to examine individual differences in the construct (Meindl & Ehrlich, 1988). Whereas the tendency for attributions towards leaders is consistent across gender, education, job tenure, and span of control, evidence suggests older individuals tend to romanticize leaders more than younger individuals (Meindl & Ehrlich, 1988).

A significant proposition of ROL theory has examined charisma as a hyper-romanticization of leadership (Meindl, 1990). In a meta-analysis of 18 studies, Schyns, Felfe, and Blank (2007) found that ROL as an individual difference variable is indeed related to follower perceptions of charisma. Further studies examined the types of leaders to which charisma and ability might be over-attributed. For example, Kulich, Ryan, and Haslam (2007) found that evidence of ROL in attribution behaviors is consistent between male and female leaders based on the relationship between organizational performance and perceived leader charisma. However, gender mattered when examining the allocation of bonuses: female leaders were rewarded less (in bonuses) for success but also punished less for failure (in bonuses). Further, Heilman and Haynes (2005) found that gender also mattered when joint performance evaluations were given to mixed-sex dyads, with attributions biased positively towards men on competence, influence, and leadership.

More recently, several studies have examined the concept of ROL in relation to other leadership concepts and outcomes. Felfe and Schyns (2014) found that a propensity towards

ROL positively relates to motivation to lead and that this relationship is stronger for individuals high in self-efficacy and personal initiative. Felfe and Petersen (2007) found that individuals high in their tendency to romanticize leaders consider information about the probability of success of the leader more than those who were lower on this tendency and were also favorable about projects with high leader success probability even when the situation was unfavorable. These effects even held when participants were told that the project's success was not dependent on the leader (Herrmann & Felfe, 2009). Carsten and Uhl-Bien (2013) found that ROL as an individual difference influenced the relationship between the co-production of leadership beliefs and displacement of responsibility (moral disengagement) such that this relationship was stronger for individuals high in ROL.

Taking these studies together, evidence suggests that: a) ROL relates to other leadership concepts, especially perceptions of charisma (Schyns et al., 2007); b) leader attributes, such as gender, matter for some outcomes (Heilman & Haynes, 2005; Kulich et al., 2007); c) individuals vary in the extent to which they over-attribute firm performance to leadership (Meindl, & Ehrlich, 1988); and d) ROL impacts decision-making (Felfe & Petersen, 2007) and responsibility (Carsten & Uhl-Bien, 2013). We believe it is important to return to the original test of the model (Meindl & Ehrlich, 1987) by examining the influence of leadership attributions on perceptions of organizational outcomes.

Replication Approach

Brandt and colleagues (2014) note that “when focusing on a theoretical prediction rather than effects within a given paradigm, a combination of close and conceptual replications is the best way to build confidence in a result” (p. 222). In this paper, we perform both close (Studies 1 & 2) and conceptual (Studies 3 & 4) replications (Schmidt, 2009). Following the design of

Meindl and Ehrlich's (1987) theory-testing, Studies 1 and 2 are close replications (Erdfelder & Ulrich, 2018; Hüffmeier, Mazei, & Schultze, 2016). In addition to testing the extent to which the original study results hold, close replications estimate the effect size expected in further replications (Brandt et al., 2014). Thus, we use Studies 1 and 2 to better estimate the effect sizes expected in Studies 3 and 4, which serve as conceptual replications (Crandall & Sherman, 2016; Schmidt, 2009) that include additional conditions intended to address criticisms regarding the use of student samples (Henry, 2008), potential gender differences in ROL effects (Kulich et al., 2007), and limited evidence regarding the attribution of failures (Bligh et al., 2011; Meindl, Ehrlich, & Dukerich, 1985).

It is important to note that decision criteria for a successful replication are different for each type of replication. In addition, we also note that a failure to replicate Study 1 or Study 2 does not mean that Studies 3 and 4 are redundant. Indeed, we argue that, while the specific study conducted by Meindl and Ehrlich (1987) may or may not be replicable, we also need to assess the generalizability of ROL to broader contexts (e.g., female versus male leaders; success vs. failure) and if the conceptual space that ROL covers (e.g., not only success but also failure) can be replicated in a further study. Below, we present each study's hypotheses, procedures, analytical techniques, and results.

CLOSE REPLICATION: STUDIES 1 & 2

The goal of Study 1 was to replicate Meindl and Ehrlich's (1987) study as closely as possible. Close replication means that design, materials, independent and dependent variables, as well as analyses, will remain the same as in the original study, although the authorship differs (Hüffmeier et al., 2016). Brandt and colleagues (2014) suggest applying the following criteria for close replication studies:

“1. Carefully defining the effects and methods that the researcher intends to replicate; 2. Following as exactly as possible the methods of the original study (including participant recruitment, instructions, stimuli, measures, procedures, and analyses); 3. Having high statistical power; 4. Making complete details about the replication available, so that interested experts can fully evaluate the replication attempt (or attempt another replication themselves); 5. Evaluating replication results, and comparing them critically to the results of the original study.” (p. 218)

Overview of Meindl and Ehrlich (1987)

The present study is intended to replicate the procedures used by Meindl and Ehrlich (1987), particularly in their Study 1, as precisely as possible to determine if the findings continue to hold more than 30 years later. Meindl and Ehrlich (1987) used experimental conditions to examine the nature and implications of a romanticized conception of leadership by manipulating the causal attributions made for organizational performance outcomes. Rather than examining the causal attributions directly (as was done in most follow-up studies of ROL), their study examined the extent to which the attribution affected subsequent reactions and perceived value of the organization. It was one of the first tests of the effects of ROL. Specifically, Meindl and Ehrlich (1987) “explored the possibility that this value is powerful enough to influence the evaluations of outcomes to which it is attached and were particularly interested in demonstrating that varying the degrees to which performance outcomes can be credibly attributed to leadership would result in measurable differences in how judges evaluated those outcomes” (p. 94). In other words, the approach taken highlights not simply to what extent individuals over glorify or vilify leaders, but to what extent it affects the perceived value of the organizations in which leadership

operates. This specific theory-testing approach has not received much research attention (see Hino & Aoki, 2013 for an exception).

The first of Meindl and Ehrlich's (1987) two studies was conducted with 111 MBA students who were asked to evaluate a firm's profitability and risk based on an informational "fact sheet" that contained data on sales, profits, and stock prices. The study also provided the manipulation of different causal attributions, emphasizing either a) leadership, b) quality of employees, c) changes in the market, or d) regulatory changes (between-subjects design). Findings highlighted that participants in the condition that emphasized leadership rated profitability higher and risk lower than the other conditions. Participants also rated firm performance more positively on a composite positivity scale (good-bad, positive-negative, attractive-unattractive, etc.). As such, we likewise hypothesized that "the value and significance ascribed to leadership as a causal force tends to enhance the subjective value of its presumed effects" (Meindl & Ehrlich, 1987, p. 94).

Replication Hypothesis 1: *Individuals presented with leadership-attributed causal accounts will evaluate firm performance more favorably in terms of a) profitability, b) risk, c) overall performance, and d) positivity than when the same performance outcomes are attributed to non-leadership factors.*

STUDY 1

In keeping with the original study, graduate business student participants were recruited through MBA programs at multiple universities in the United States. To assess the minimum required sample size, we conducted an *a priori* power analysis in G*Power 3.1.9.2 (Faul, Erdfelder, Buchner, & Lang, 2009). Assuming an effect size f (calculated from $\eta^2 = .056$ from the original study, see below) of .24, we determined that a sample size of 376 participants total

would be adequate to obtain 95% power to detect our hypothesized effects at $\alpha = .0125$ (α of .05 divided by the 4 outcomes variables) with 4 groups. Thus, we planned to collect data from a minimum of 400 participants to ensure an adequate sample size. As in the original study, we used a between-subjects design—that is, each student was randomly assigned to one of four online experimental conditions (i.e., leader, quality of employees, market trends, and government regulations).

Procedures

As in the original study, participants were given written instructions explaining that the purpose of the questionnaire was to examine the effectiveness of certain business and financial information in allowing people to form general impressions and evaluations of firms traded on the New York Stock Exchange. As such, the participants were asked to evaluate one firm based on information from several business journals and investment references condensed into an official-looking fact sheet (Meindl & Ehrlich, 1987, p. 96). Because the fact sheet used in the original study was not published, we developed a new fact sheet (available in Appendix A) based on the description in the original study. The fact sheet includes an overview of a fictitious pharmaceutical company, ACTRON, using the same categories of data that were presented in the original study (Meindl & Ehrlich, 1987, p. 97): 1) a description of the firm's key operating strengths, and 2) performance indicators from the past five years (total sales, profit margins, net earnings, earnings per share, and the company stock price).

In the original study, “the content of the paragraph describing key operating strengths was manipulated to create four distinctly different causal accounts of the firm's performance in recent years” (Meindl & Ehrlich, 1987, p. 97). The conditions attributed the firm's performance to 1) the leadership of the company, 2) the quality of the employees, 3) market trends, and 4)

governmental regulation. These causal accounts were published in the original study and we used them in the same manner. They can also be found in Appendix A.

To ensure the quality of our data, we included two attention check questions using an agreement Likert scale (“I answered the questions carefully” and “Please respond ‘2’”). We also screened for very short survey duration and no standard deviation within the scales, indicators of a careless response style.

Sample

Overall, we contacted 25 business schools across the USA. Of those, six schools declined our request to send the questionnaire to their students. Thus, the questionnaire was sent out to 19 business schools. In the survey, we received 396 clicks of which 10 were not MBA students, and 88 did not complete the survey. Therefore, 298 MBA students from 14 different US business schools completed the survey in its entirety. We subsequently conducted several quality checks. First, we deleted those that did not select “2” on the slider scale as instructed ($N = 6$). Eleven participants answered “strongly disagree” to “how carefully I answered the questions” yet met all the other inclusion criteria. As the intent for this question was to remind participants to be more mindful, we decided to keep these participants in the sample after conducting a t-test between those 11 and the full sample and finding no differences in the outcome variables. No further participants were deleted due to a short completion time or zero standard deviation in answering the questionnaire. Thus, the final sample size was 292. While this fell short of our intended sample size based on the power analysis we conducted, it is much larger than Meindl and Ehrlich’s (1987) original sample ($N=111$).

Of the participants who provided demographic information, 157 were male and 127 were female (7 preferred not to answer and 1 preferred to self-identify). The mean age was 31.37 years

(SD = 7.95), thus slightly older than the original sample (25 years old). Of the participants, 175 worked full-time, 61 worked part-time and 56 were not employed. On average, they worked 34.55 hours a week (SD = 16.05) and had 6.09 years of work experience (SD = 6.63). The sample size per condition was as follows: government (N = 75), market (N = 70), employees (N = 73), and leadership (N = 74).

Measures

Firm performance evaluations were rated by four indicators as in Meindl and Ehrlich (1987): profitability, risk, overall performance, and positivity.

Profitability and risk. After reading the fact sheet, participants were asked to evaluate the firm's performance. As in the original study, the participants were first asked to rate dimensions of profitability and risk associated with the organization using two separate items, each with a 10-point response scale ranging from 0 (not at all profitable/risky) to 10 (very profitable/risky).

Overall performance. Using the same procedure as Meindl and Ehrlich (1987), we created a single measure of positive overall performance by combining high profit and low risk (reverse-coded from the survey presentation).

Semantic differential (Positivity). Following the ratings of profitability and risk, participants were asked to rate the firm's performance on a measure of "positivity" using nine bipolar, 11-point semantic differential scales: good-bad, unstable-stable, successful-unsuccessful, risky-safe, positive-negative, variable-constant, profitable-unprofitable, uncertain-certain, and attractive-unattractive. In line with Meindl and Ehrlich (1987), we combined the nine semantic differential items into a single positivity scale. The reliability for this scale was $\alpha = .84$, slightly lower than Meindl and Ehrlich (1987; $\alpha = .88$).

Demographic information: At the end of the survey, participants were asked for demographic information including age, work experience, and gender. In this study, we controlled for age and work experience. Meindl and Ehrlich (1988) found that age is related to ROL, with older participants showing higher ROL. In addition, we controlled for work experience to account for any changes in work experience requirements and differences between programs for MBA students over the last thirty years.

Table 1 depicts the means, standard deviations, and intercorrelations among the control variables, experimental conditions, and all dependent variables.

Insert Table 1 about here

Analysis

Following the analytical techniques used by Meindl and Ehrlich (1987), we conducted a multivariate analysis of variance (MANOVA) along with planned comparisons using the Bonferroni correction for the dependent variables of profitability, risk, and overall performance, as well as positivity using SPSS version 26. We also conducted a MANCOVA with age and work experience as control variables.

Results

We expected that evaluations of the company in the leader condition would differ significantly from the other conditions (market, employees, and government) for profitability, risk, performance, and the semantic differential. Following a conservative test with adjusted alpha levels of .0125 per test ($.05/4$), we found no significant differences across any of the

conditions with or without the inclusion of control variables. These results can be found in Table 2.

Insert Table 2 about here

As Meindl and Ehrlich (1987) used alpha levels of .05 as their test of significance, we also considered a non-adjusted level of significance. With this consideration, the analyses showed significant differences between the conditions for the semantic differential [$F(3, 287) = 3.05, p = .03$, see Table 2]. Further exploration showed the largest paired comparison was between government ($M = 6.73$) and leadership conditions ($M = 7.30$), which was not significant ($p = .05$). When including control variables, the main effect of condition was slightly weaker [$F(3, 269) = 2.73, p = .04$] and the significance for this paired comparisons became even weaker at $p = .09$. We did also find a significant intergroup difference for risk. This was weakly significant in the MANOVA [$F(3, 287) = 2.57, p = .05$], yet was stronger after including the control variables (age and work experience) [$F(3, 269) = 2.86, p = .04$]; however, the most notable group difference found in the post-hoc comparison was between the employee and the government conditions ($p = .05$), which we did not hypothesize.

We present results of the MAN(C)OVAs with the unrestricted sample as well as multiple regressions using dummy variables against the leader condition in an online supplementary file. The results are very similar when including the six excluded participants in overall effects, with the only significant planned comparisons emerging for employee and government conditions for risk in the MANCOVA. Likewise, regression results revealed a similar overall pattern with the government condition reaching marginal significance (against the leadership reference) for risk

and the semantic differential when including control variables. Taken together, we did not find robust evidence that firms were evaluated more positively when presented with leadership-attributed causal accounts than other explanations.

STUDY 2

Since the Study 1 sample remained somewhat lower than expected, despite our efforts to collect a large sample of MBA students in line with our initial power analysis; we collected an additional, more heterogeneous sample to again test Hypothesis 1. Specifically, we collected a sample of working adults from the United States recruited using a panel research provider: Respondi (www.respondi.com). The design, procedure, and instruments were the same as those of Study 1. However, Respondi exits participants from the survey if they do not pass the attention checks, which in our case meant that for those participants, we do not have demographic data.

Sample and instruments

Overall, we collected data from 596 participants. We did not include responses from 39 who did not consent to take part, 37 who did not pass the attention checks, and 25 who had zero standard deviation. A further 47 participants were deleted due to short survey duration (one standard deviation below the median duration; less than 90 seconds) in answering the questionnaire, leaving a final useable sample of 448. Respondents were paid \$2.13 for their participation.

Of those participants who provided demographic information, 254 were male and 193 were female (1 preferred not to answer). The mean age was 45.45 years ($SD = 10.74$). Seven worked part-time and 439 worked full-time, whereas two were not currently working. The average hours worked per week was 42.11 ($SD = 7.03$). The average work experience was 14.36

years ($SD = 10.24$). The sample size per condition was as follows: government ($N = 113$), market ($N = 112$), employees ($N = 110$), and leadership ($N = 113$).

We used the same measures as in Study 1. The reliability for the semantic differential in this sample was $\alpha = .91$ (Meindl & Ehrlich, 1987; $\alpha = .88$).

Insert Table 3 about here

Analysis

Table 3 depicts the means, standard deviations, and intercorrelations between the control variables, experimental conditions, and all dependent variables. Again, we used a multivariate analysis of variance (MANOVA) along with Bonferroni-adjusted planned comparisons for the dependent variables of profitability, risk, and overall performance, as well as positivity using SPSS version 26. We also conducted a MANCOVA with age and work experience as control variables. Results are presented in Table 4.

Results

As before, we first examined the differences between the conditions without control variables. None of the differences were significant: profit [$F(3, 444) = .63, p = .60$], risk [$F(3, 444) = .38, p = .77$], performance [$F(3, 444) = .53, p = .66$], and semantic differential [$F(3, 444) = 1.74, p = .16$ see Table 4]; neither were any of the post-hoc paired comparisons. Including the control variables (age and work experience) equally did not lead to any significant differences between the conditions. None of the post-hoc comparisons were significant. Therefore, Hypothesis 1 was not supported in Study 2. The results remained essentially the same when including the excluded participants. Additionally, we present results of the MAN(C)OVAs with the unrestricted sample and multiple regressions in an online supplementary file.

Insert Table 4 about here

In addition, we combined the samples of Study 1 and Study 2 (See Table 5). Controlling for sample, we found no significant effects for profitability, risk, or overall performance. Only for the semantic differential was the condition effect significant [$F(3, 734) = 4.18, p = .01$]. Post-hoc analyses showed significant mean differences between the leadership ($M = 7.38$) and government ($M = 6.88$) conditions as well as the government ($M = 6.88$) and employee ($M = 7.31$) conditions with the leadership condition being evaluated most favorably. The pattern of relationships was similar when including age and work experience as controls; however, when doing so, the post-hoc comparison between government and employee conditions was no longer significant.

Criteria for a Successful Close Replication

The aim of Studies 1 and 2 was to conduct a close replication of Meindl and Ehrlich's Study 1. In order to determine the success of the replication, we first looked at the significance of our results. Statistically significant differences between the leadership condition and non-leadership conditions in the same direction as Meindl and Ehrlich's results would have indicated a successful replication. However, no consistent post-hoc comparisons emerged in Studies 1 and 2, apart from one effect between government and leadership for the semantic differential that only approached significance within Study 1 and was significant in the combined data. We then compared the effect sizes, considering Cohen's (1988) rules of thumb for small, medium, and large effects, expecting them to be in the same range. Although Meindl and Ehrlich did not

include effect sizes, we calculated their effect sizes based on the results reported in Table 1 of their article. We used the following formula:

$$F(df1, df2) = \eta^2 / (1 - \eta^2) \times df2 / df1,$$

where η^2 is the standardized effect size (variance explained in the DV by group differences). For profitability, risk, and overall performance this means:

$$F(3,107) = \eta^2 / (1 - \eta^2) \times 107 / 3$$

Solving the solution for η^2 leads to:

$$\eta^2 = 1 / (1 / (F \times 3 / 107) + 1).$$

Following this calculation, the effect sizes for Meindl & Ehrlich's data are as follows: profitability, $\eta^2 = 0.092$, risk, $\eta^2 = 0.056$, overall performance, $\eta^2 = 0.120$, and semantic differential, $\eta^2 = 0.074$. We present effect sizes and F-tests comparing results of Meindl and Ehrlich (1987) with Study 1 and Study 2 and combined samples in Table 5. For all outcomes, our effect sizes were considerably smaller than those of the original study and the post-hoc test did not show the expected mean differences. We, therefore, conclude that Study 1 and Study 2 do not robustly replicate the original study by Meindl and Ehrlich.

Insert Table 5 about here

CONCEPTUAL REPLICATION: STUDIES 3 & 4

The goal of Studies 1 and 2, close replication, was to replicate as closely as possible Meindl and Ehrlich's (1987) study. The intent of our conceptual replications is to investigate the replicability of the concept of ROL more broadly. We argue, along with Crandall and Sherman (2016), that conceptual replications can move the field forward towards a better generalization of

a theory, considering different operationalizations. As such, even though Studies 1 and 2 did not replicate Meindl and Ehrlich's results, the theory of ROL may still be valid and thus warrant a conceptual replication. According to Schmidt (2009), conceptual replications are riskier than close replications, as a failure to replicate can be explained by different reasons such as sampling error or misconceptions of the new experimental setup. However, if a conceptual replication is successful, it contributes to supporting a theory (Schmidt, 2009), rather than merely a specific experiment. We argue that conceptual replication can be the starting point of further research into understanding under which circumstances ROL applies.

Studies 3 and 4 address limitations in the original experiment and account for developments in the study of ROL since the publication of Meindl and Ehrlich's (1987) work. The conceptual replication was conducted in two parts (Studies 3 and 4) and uses Meindl and Ehrlich (1987) as a template, with a few modifications. Both parts of the study include a) a broader sample, b) a failure condition, c) a female leader condition, and Study 4 includes d) additional outcome variables that indicate behavioral intentions.

First, we included a broader sample. Meindl and Ehrlich (1987) note in their discussion of the original research: "Although most of our subjects held full-time positions in organizations and thus were not typical full-time undergraduates, they hardly represent the range of possible organizational constituencies and stakeholders" (p. 106). That is, there is a possibility that the results obtained are due to lack of work experience or a young age. The question is whether, and to what extent, older individuals with more experience will romanticize leaders in the same way. There is some evidence that ROL is stronger for older than younger individuals (Meindl & Ehrlich, 1988). Additionally, using MBA students will necessarily narrow the work background of the sample to those that feel they need further studies for their career progress. Likewise,

Henry (2008) suggests replication as a solution to the use of student samples. Thus, in Studies 3 and 4, we used samples of working adults instead of graduate students. Although we did not attempt to address the age range of all possible stakeholders in Studies 3 and 4, we extended the samples to include participants with a broader range of ages and work experiences. This allowed us to investigate the extent to which ROL may generalize to samples that are more heterogeneous.

Second, whereas Meindl and Ehrlich's (1987) study only included organizational success, there is a precedent for the impact of positive or negative performance (Meindl et al., 1985). Overall, the ROL theory suggests that leadership attributions may be impactful both in times of success and failure (Meindl, et al., 1985). Bligh, Kohles, Pearce, Justin, and Stovall (2007) also found a tendency to over-attribute blame for negative follower outcomes to leadership. In a study of undergraduate students in Japan, Hino and Aoki (2013) found a stronger effect for a leader condition than an external condition (but not an employee condition) in conditions of organizational failure. Gibson and Schroeder (2003) conducted a study examining attributed credit and blame to organizational leaders. Their findings highlighted that, whereas observers assigned *credit* more equally throughout hierarchical levels in organizations; *blame* tended to rise in hierarchical levels—that is, they blamed upper-level leaders to a greater degree than lower-level leaders. This could imply that upper-level leaders are blamed even more for failure than praised for success, meaning that ROL might be even stronger for failure than success. Also, observers attributed more credit and blame to individuals rather than groups, showing the relevance of ROL.

Third, we included a female leader condition. Although romanticizing leadership may be attributable to the situation (success/failure) and the observer (the sample), ROL may also

depend on the characteristics of the leader. More specifically, previous research has suggested that individuals attribute different degrees of influence on leaders who are perceived as more prototypical (Shamir, 1992). The original studies by Meindl and Ehrlich only used a male leader in their vignettes, as did many follow-up studies (Felfe & Petersen 2007, Schyns & Hansbrough, 2012). Very little research has examined ROL for female leaders, despite women holding 51.5% of managerial roles in the United States (U.S. Department of Labor, 2017). To extend the work on gender and ROL, and in response to the call from Bligh and Schyns (2007) to examine the extent to which leader characteristics affect ROL, we include a female leader condition. This allows us to explore whether the specific attributional process of ROL is limited to male leaders (i.e., a more prototypical gender role) alone or if it generalizes to female leaders.

Kulich and colleagues (2007) suggest four reasons why ROL may be less pronounced for female leaders. First, performance evaluations are different for men and women such that women are generally evaluated less favorably. Second, leader prototypical traits are more strongly associated with stereotypically male attributes and therefore women may receive less “credit” for achievement. Third, gender stereotypes for men are more agentic suggesting that men may be more likely to be viewed as a source of change and progress than women. Finally, women tend to have lower group status than men, which is associated with instrumental outcomes and influence. Findings from Kulich and colleagues suggest an overall romance for both men and women, but there is variability “underneath the surface” such as differences in how ROL effects benefit men in terms of pay or charismatic attributions (both higher for the male condition). Given other studies (particularly in laboratory settings) regarding gender differences in evaluations of leadership (Eagly & Johnson, 1990; Eagly, Karau, & Makhijani, 1995; Heilman & Haynes, 2005; Kulich et al., 2007), we expect that perceived organizational evaluations will be

lower for the female leader than for the male leader, but nonetheless stronger than non-leadership attributions, particularly in instances of organizational success.

Despite these expectations that in situations of organizational success male leaders will receive more credit, in situations of organizational failure, women may receive more blame. For example, even a single performance failure may signal a lack of competence for women, especially as perceived to be in a role inconsistent with their gender (Brescoll, Dawson, & Uhlmann, 2010). Kulich and colleagues (2007) hypothesize a stronger penalty for female CEOs in times of failure; however, they found that in terms of bonuses, female CEOs received less penalty for poor organizational performance than men. Whereas female leaders might not be penalized financially as much as male leaders, we hypothesize that organizations will receive more negative evaluations in times of failure when the failure is attributed to a female CEO.

Park and Westphal (2013) suggest female and minority CEOs are likely to be blamed for poor organizational performance in the media. Through archival data, interviews, and surveys with CEOs and journalists, they found evidence that white male CEOs attributed low performance in organizations to a female or racial minority CEO in conversations with journalists. That is, when commenting on other firms' performance, they were more likely to attribute the low performance internally when the CEO was a female or minority. Subsequently, white male CEOs' internal attributions increased the likelihood that white male journalists reported the "blame." The authors suggest that out-group biases and status competition elicit negative forms of envy toward higher-status minority CEOs, and this attribution is translated into reporting in the media. The public, in turn, likely hears and perhaps internalizes these explanations, which may manifest as an increased devaluing of the company when poor performance is attributed to a female leader in contrast to a male leader or other outcomes.

We thus expect an interaction between success/failure and leader gender.

Hypothesis 2a: *In success conditions, firm performance evaluations will be rated more positively in the male leader-conditions than female leader-conditions, and both of which are more positive than market conditions.*

Hypothesis 2b: *In failure conditions, firm performance evaluations will be rated more negatively for the female leader-conditions than male leader-conditions, both of which are more negative than market conditions.*

STUDY 3

Procedures

We conducted a study to examine the conceptual replicability of Meindl and Ehrlich's (1987) results. Here, we collected an online sample comparing six conditions based on a 3x2 design: causal account (male leader, female leader, and market forces) and firm performance (success vs. failure). Participants were asked to complete a short demographic survey.

Participants completed similar procedures to those conducted in Study 1 (e.g., presentation of a fact sheet, presentation of one of the six causal account conditions, and ratings of profitability, risk, and positivity).

We used a 3x2 study design with the causal account (male leader, female leader, and market forces) and firm performance (success vs. failure) as treatment conditions (see Appendix B for Study 3 and 4 materials). Effectively, this means that we compared six conditions with each other, namely, male leader success, male leader failure, female leader success, female leader failure, market success, and market failure. To increase the power in the study, we reduced the alternative conditions from Studies 1 and 2 to focus only on the market condition, which was found to be the most different from the leader condition by Meindl and Ehrlich

(1987). To assess the minimum required sample size, we conducted an *a priori* power analysis in G*Power 3.1.9.2 (Faul et al., 2009). Assuming an effect size f (calculated from $\eta^2 = .056$ from the original study) of .24, an α of .00625 ($\alpha = .05$ divided by the number of outcome variables), a power of .95, 7 degrees of freedom, 6 groups, and 2 control variables, we arrived at a final necessary sample size of 530.

Sample

Overall, we collected data from 648 participants in the United States using a panel research provider: Respondi (www.respondi.com). We did not include responses from 44 who did not pass the attention checks and 13 who had zero standard deviation. A further two participants were deleted due to short survey duration, leaving a final useable sample of 589. Participants were paid \$2.13 for participation in this study.

Of those participants, 364 were male, 323 were female, and 2 respondents prefer to self-identify. The mean age was 41.35 years ($SD = 11.85$). On average, they worked 41.5 hours a week ($SD = 6.62$) and had 9.4 years of tenure ($SD = 8.68$). The sample size per condition was as follows: market failure ($N = 100$), market success ($N = 97$), female leader failure ($N = 94$), male leader failure ($N = 102$), female leader success ($N = 103$), and male leader success ($N = 93$).

Measures

Firm Performance. Firm performance evaluations were assessed using the same measures for participant ratings of profitability, risk, overall performance, and positivity as in Studies 1 and 2. The reliability for the nine items of the semantic differential was $\alpha = .95$.

Control variables. We used tenure and age as control variables as age has been found to influence ROL (Meindl & Ehrlich, 1988). Table 6 depicts the means, standard deviations, and

intercorrelations among the control variables, experimental conditions, and all dependent variables.

Insert Table 6 about here

Analysis

To test Hypothesis 2a and 2b, we conducted a MANOVA with the three causal accounts (male leader, female leader, and market) and success/failure conditions as treatment conditions as well as their interaction along with planned comparisons using the Bonferroni correction. We also conducted a MANCOVA including age and tenure as control variables. All analyses were conducted within SPSS version 26.

Results

We expected that the leader conditions would differ significantly from the market conditions and that there would be significant differences between the male and female leader conditions with respect to the outcome variables of profitability, risk, performance, and the semantic differential. Table 7a presents the results of these analyses and Table 7b indicates the means for each condition. For all dependent variables only the success/failure condition was significant: profitability [$F(1, 583) = 361.86, p < 0.001$]; risk [$F(1, 583) = 73.79, p < 0.001$]; performance [$F(1, 583) = 361.65, p < 0.001$]; and semantic differentials [$F(1, 583) = 467.28, p < 0.001$], indicating a more positive response in success conditions. Neither causal account (male leader, female leader, and market conditions) nor the interaction was significant. No hypothesized differences between male and female leadership in success and failure conditions emerged, nor did differences between leadership and market conditions. The results remained the

same after controlling for age and tenure. As such, Hypothesis 2a and 2b were not supported.

Additionally, we present results of the MAN(C)OVAs with the unrestricted sample and multiple regressions in an online supplementary file with similar results. The results remained essentially the same when including the excluded participants.

Considering that the results could be driven by a short response time, (i.e., that participants who completed the survey in under 2 minutes had not properly understood or paid close enough attention to the vignettes), we repeated the analyses selecting only participants who took more than 2 minutes to fill in the questionnaire. The results remained the same.

Insert Table 7a & 7b about here

Discussion

The aim of Study 3 was to replicate the results obtained by Meindl and Ehrlich (1987) with two additional conditions. Specifically, we added the gender of the leader (male/female) and success versus failure. In the original study, only a male leader was used and only a success condition. Contrary to our expectations, we found no differences between the gender conditions or the leader versus market conditions. The only significant difference regarding our outcome measure was due to the success versus failure conditions, which is to be expected and was not a hypothesized difference. The effect sizes for hypothesized effects were nearly zero (ranging from $\eta^2 = .00 - .01$). Therefore, we conclude that Study 3 did not conceptually replicate the results of Meindl and Ehrlich's (1987) original study.

STUDY 4

We used the results of Study 3 to further consider our design for Study 4. We considered that due to the online nature of the presentation of the company description, participants may not

have paid enough attention when reading the descriptions and simply clicked “next” to answer the questions without reading closely. So, for Study 4, we changed the online survey settings such that participants could not continue through the survey until after 60 seconds of exposure to the company description to encourage participants to read the text more carefully.

Further, to extend and add to our understanding of the effects of ROL, we also included additional outcome measures over and above those suggested by Meindl and Ehrlich (1987). Here, we wanted to examine how far the ROL also generalizes to other outcomes. In addition to considering participants’ evaluations of the hypothetical company, we also wanted to understand the outcome of ROL on behavioral intentions. Borrowing from Kim (2014), we examine behavioral intentions as activities that any individual might perform or consider performing with a company. These behavioral intentions constitute a personal stake in the company, rather than generalized evaluations of the organization including the individual’s intent to support, invest in, seek employment with, or purchase from the company. Therefore, extending outcome variables from Study 3, we hypothesize the following:

Hypothesis 3a: *In success conditions, firm performance evaluations and behavioral intentions will be rated more positively in the male leader-conditions than female leader-conditions, and both of which are more positive than market conditions.*

Hypothesis 3b: *In failure conditions, firm performance evaluations and behavioral intentions will be rated more negatively for the female leader-conditions than male leader-conditions, both of which are more negative than market conditions.*

Sample

Again, we compared six conditions, namely, male leader success, male leader failure, female leader success, female leader failure, market success, and market failure. As in Study 3, a

sample from the United States was recruited using a panel research provider: Respondi (www.respondi.com). Participants were paid \$2.13 for participation in this study. We used the same *a priori* power analysis as in Study 3 with a final necessary sample size of 530.

We initially collected responses from 633 participants. Twenty-eight participants were deleted as they failed the attention check questions. As before, 11 participants were deleted based on zero standard deviations. In this sample, no respondents were deleted as we looked at the survey duration. Therefore, our final sample size was 594. The distribution into the conditions was as follows: male leader success (N = 101), male leader failure (N = 89), female leader success (N = 105), female leader failure (N = 101), market success (N = 114), and market failure (N = 84).

Of the participants, 382 were male and 208 were female (4 respondents chose not to answer or prefer to self-identify). The average age was 44.78 years old (SD = 11.34). The average work experience was 18.18 years (SD = 11.46) and the average tenure was 10.30 years (SD = 8.89).

Procedures

All study activities were conducted online. Participants were asked to complete a demographic survey before they participated in the experiment. Similar to Studies 1, 2, and 3, participants were presented with a fact sheet, were presented one of six conditions (see above), and ratings of profitability, risk, and positivity. After the participants completed their ratings, they also completed a short behavioral survey to gauge their intent to support, seek employment with, invest in, and purchase from ACTRON, the fictitious company in the scenarios (Kim, 2014). By forcing participants to stay on the page with the time lag feature along with the attention check items, we tried to ensure participants were attentive to the scenarios.

Measures

Firm Performance. Firm performance evaluations were assessed using the same measures for participant ratings of profitability, risk, overall performance, and positivity in Study 4 as in the previous 3 studies. The reliability of the semantic differential was $\alpha = .93$.

Behavioral Intentions. Twelve items were used to capture the behavioral intentions of the participants toward the fictional firm presented in the scenarios (Kim, 2014). Three items focus on intent to support the company (e.g., *I would talk positively about the company with others*; $\alpha = .92$). Three items examine intent to invest in the company (e.g., *I think the company is a good company to invest in*; $\alpha = .93$). Three items examine intent to seek employment with the company (e.g., *I would like to seek employment opportunities with the company*; $\alpha = .90$). And finally, intent to purchase from the company is assessed using three items (e.g., *I would like to buy products from the company*; $\alpha = .93$).

Control variables. We used years of work experience and age as control variables. Table 8 depicts the means, standard deviations, and intercorrelations of the control variables, experimental conditions, and all dependent variables.

Insert Table 8 about here

Analysis

To test Hypothesis 3a and 3b, we conducted a MANOVA with causal account (male leader, female leader, and market forces) and firm performance (success vs. failure) as treatment conditions as well as their interaction with the eight outcomes as the dependent variables and compared the groups using Bonferroni adjusted pairwise comparisons using SPSS version 26.

Additionally, we conducted all analyses with the inclusion of the control variables, using a MANCOVA.

Results

MAN(C)OVA results are presented in Table 9a, and Table 9b indicates the means for each condition. For all dependent variables, only the success/failure condition was significant: profitability [F (1, 583) = 284.66, $p < 0.001$]; risk [F(1, 583) = 52.93, $p < 0.001$]; performance [F (1, 583) = 290.02, $p < 0.001$]; semantic differential [F (1, 583) = 411.05, $p < 0.001$]; support intentions [F (1, 583) = 214.08, $p < 0.001$]; investment intentions [F (1, 583) = 322.12, $p < 0.001$]; employment intentions [F (1, 583) = 182.83, $p < 0.001$]; and purchase intentions [F (1, 583) = 131.74, $p < 0.001$]. Neither causal account (male leader, female leader, and market conditions) nor the interaction of casual account and failure condition were significant for any outcomes. The pattern of MANCOVA results was the same for all dependent variables except for profitability in which the interaction effect was significant [F (2, 581) = 3.31, $p = .04$], but none of the paired comparisons were significant. Similar to Study 3, hypothesized effect sizes were very small, ranging from $\eta^2 = .00 - .01$. As such, neither Hypothesis 3a nor 3b was supported in Study 4. The results remained essentially the same in the unrestricted sample, with the exception that the interaction effect was also significant for the semantic differential and causal account was significant for support intentions, both only while including controls, indicating marginal effects. We present results of the MAN(C)OVAs with the unrestricted sample and multiple regressions in an online supplementary file.

Insert Tables 9a and 9b about here

Criteria for a Successful Conceptual Replication

Like our criteria for Studies 1 and 2, our criteria for successful conceptual replication was based on both statistical significance and effect size. We considered our conceptual replication successful if the significance of the effects and effect sizes for Hypotheses 2 and 3 were comparable to our Studies 1 and 2 for the leadership conditions compared to market conditions. ROL theory is “gender-neutral,” as the theory only discusses leaders generally; therefore, we expected differences for both male and female leaders against the market conditions. In addition, we considered the conceptual replication successful if the effect sizes for the failure and success conditions are in the same range using Cohen's (1988) rules of thumb for small, medium, and large effects. Similar to all three previous studies, the expected differences did not emerge, and the effect sizes were extremely low, even lower than Meindl and Ehrlich's effect sizes, within the success and failure conditions. Taken together, we conclude that we could not replicate the ROL study by Meindl and Ehrlich (1987), nor could we show that the ROL generalizes to more heterogeneous samples or behavioral intention outcomes. We also could not show that ROL extends to the gender of the leader and situations of success and failure using this design.

GENERAL DISCUSSION

Our series of four replication studies aimed to re-examine early claims of the Romance of Leadership theory by replicating Meindl and Ehrlich's (1987) Study 1. Specifically, these studies investigated the impact of leadership attributions on the evaluation of organizational performance. We found it interesting that although the study of ROL has resulted in many offshoots, the original ideas based on the relationship between ROL and performance evaluations had not been replicated with the exception of Hino and Aoki (2013). In the first two of our studies, we conducted close replications, that is, we kept the design (Study 1 and Study 2) and

sample (Study 1) as similar as possible to the original study. We conducted two further studies that expanded the design of the original study by including a female leader and a failure condition. Although the original study only focused on the effect of success and only used a male leader in the company description, we wanted to examine how far the design would extend to the theoretically assumed failure effects (Meindl et al., 1985) and if the ROL phenomenon would extend to female leaders (Kulich et al., 2007). In our final study, we captured additional outcome variables to examine behavioral intentions beyond general evaluations of the company as in the original study.

Across all four studies, we found very little evidence to suggest a successful replication. When using adjusted alpha levels (.0125/.00625 per test), we could not replicate any of the results found in the original study. When using alpha levels of .05, as a direct replication of what Meindl and Ehrlich did, only one group comparison without controls approached significance for one relationship in Study 1, that was the difference between government and leadership on one of the outcomes—the semantic differential. The company was evaluated more positively in the leadership casual account than government account. Although we did not find this effect in Study 2, it was significant when combining Studies 1 and 2 with a very large sample size. Taken together, we could not robustly replicate the results obtained by Meindl and Ehrlich (1987) in our close replication (Study 1 and Study 2). The conceptual replications (Studies 3 and 4) also did not replicate Meindl and Ehrlich's findings. Further, by taking a more appropriate and conservative alpha level to avoid Type I error inflation, we found no replication at all. However, we should note that alpha adjustments may have not been common practice in 1987. For example, in his 1991 publication, Keppel noted that disregarding the increase in familywise error rate associated with the tests in factorial ANOVA was a common, yet specious practice. Taken

across the four studies, we did not find robust evidence of expected differences between leadership and other conditions. We can only speculate why we could not replicate each of the original results obtained by Meindl and Ehrlich (1987).

First, in contrast to the year 1987, more information is currently available through various media on companies and their performance, including quantitative data such as stock market valuations and qualitative examinations of what leads to good or bad performance. This could potentially have weakened the ROL phenomenon vis-à-vis the market and employee conditions, simply because participants are better informed at reading financial data and are using fewer cognitive shortcuts. The rise of evidence-based management (Rousseau, 2006), business analytics, and “big data” in MBA curricula, especially early in the MBA curricula (Passarelli, Boyatzis, & Wei, 2018), may have primed students, in particular, but the sample from each study in general, towards a value of data in decision-making and evaluations. This may have encouraged participants to weigh more heavily the numeric performance indicators than the narrative explanations given in the text in their evaluations of the company.

Furthermore, the trend towards big data may be coupled with a movement away from following governmental policy and civic engagement, which may explain the difference between the leadership and government account in Study 1 and the combined sample. This difference raises interesting research questions about the role of forces internal to the organization (leadership and employees) and accounts external to the organization (market and government). For example, in 2019, government and regulatory bodies may appear to be in degrees a less predictable externality than market trends. Although market trends are external to an organization, on the one hand, such trends are also driven by the organizations embedded in the

market. Perhaps capitalizing on changes in governmental regulations is clearly an external attribution for increases to the bottom line, but not viewed as “success.”

Second, Meindl and Ehrlich (1987) conducted their study using a paper-pencil format as part of an MBA program, whereas our four studies were conducted via online survey software, and three of the four studies utilized an online recruiting tool (Studies 2-4). We do not know exactly how Meindl and Ehrlich conducted the study. If, for example, data were conducted during a single class (e.g., a specific leadership-related course), this could have motivated their participants to think about leadership in terms of a demand characteristic (Orne, 1962). Instead, we chose to sample MBAs from a variety of classes, MBA programs, instructors, and universities. As such, absent from the current studies is the possibility that a single course’s content might have primed participants or made salient the role of leaders. This demand characteristic might even be stronger in a paper-pencil format handed to an instructor than completed anonymously online. It is also important to note that in all four of our studies, the average age was older than the 1987 Meindl & Ehrlich study. As age has had a higher relationship with ROL in past studies (Meindl & Ehrlich, 1988), we would have expected to see even stronger effects within an older group; however, this was not the case.

Third, the emergence of recent social movements that seem “leaderless” might also suppress the romanticizing of leadership. Events such as #MeToo, the Hong Kong protests, Black Lives Matter, Arab Spring, and Occupy Wall Street have been influential for our time, but also emerged and thrived without a leader as the driving force. Likewise, widespread media coverage of corruption and other unethical behavior by senior leaders likely increases the general distrust and cynicism towards leaders, generating the emergence of organic and leaderless protest groups made up largely of students who question the wisdom and decision-making of leaders.

Such social movements and events could quash the tendency to romanticize formal leadership vis-à-vis feeling leadership can arise from a collective.

Fourth, we introduced a female leader condition in Studies 3 and 4. We hypothesized that organizations would be given more negative evaluations in times of failure for female than male leaders. Based on Kulich et al. (2007), we expected female leaders would be evaluated less favorably than male leaders. However, we found no gender differences between the leadership conditions, nor with the market condition in either study. In several ways, this may be an encouraging finding. For example, it might indicate a reduction of explicit discrimination through exposure to an increasing number of women in leadership positions (Pew Research Center, 2015). However, it could also mean that participants were more aware of gender biases regarding women in leadership roles and answered in a socially desirable way (see examples regarding gender and political leaders: Seltzer, 2018; Streb, Burrell, Frederick, & Genovese, 2008).

Fifth, part of our failure to replicate may lie within the design limitations of the study. For example, the presentation of company performance (success and failure), may not have been strong enough to trigger a Romance of Leadership. Meindl and colleagues (1985) suggested that romanticizing is stronger in *extreme* situations of success or failure, notably, when success or failure is unexpected and, thus, requires an explanation. Perhaps the presentation of moderate success or failure did not require an explanation and such participants paid little attention to the explanation provided, choosing instead to only review the performance data; or participants may not have paid attention to the explanation for company performance due to a generalized reduction in human attention spans in the past decades due to increasing information flows (Lorenz-Spreen, Mønsted, Hövel, & Lehmann, 2019). However, even with the alteration to

Study 4 where we included a time lag on the company descriptions screen, our results remained non-significant. Alternatively, participants may be more acclimatized to fluctuations in company performance after the recessions of 2001 and 2008 and the economic recoveries that followed. Our data were collected before the global coronavirus pandemic of 2020, but we do think it would be particularly interesting to study this phenomenon *during* such a global shock.

Finally, changes in the landscape of leadership scholarship and public dialogue on the topic may have affected the replicability of Meindl and Ehrlich's 1987 study. Day and Harrison (2007) suggest the definition of leadership has evolved, moving away from an individual "leader" to considering leadership as a shared process. Likewise, research on theories such as shared and distributed leadership emerged in the late-1990s in both specialized leadership and more mainstream management journals (Ulhøi & Müller, 2014). As the leadership attribution in the present study is framed as the work of an individual leader (CEO)—not leadership as a shared process—this focus might be less relevant to today's process and shared definitions.

Taken together, we could not replicate Meindl and Ehrlich's (1987) study across four studies and over 1,800 participants. Our findings call into question the current relevance of ROL theory as indicated in the effect of leadership attributions on company evaluations. We simply did not find that leadership attributions affected evaluations of the company in conditions of success or failure regardless of the gender of the leader, and likewise did not affect behavioral intentions towards the company. In viewing ROL as an attributional bias, we might consider a failure to find evidence of this bias as good news. Perhaps the prevalence of this bias has indeed decreased due to changes in society, technology, and education. Should we consider the ROL phenomenon in general as disproven? We think this would be an over-interpretation of our results as there are several different approaches in studying the ROL phenomenon, including

measuring it as an individual difference variable or asking participants for explanations of company performance (Felfe & Petersen, 2007; Meindl & Ehrlich, 1988; Schyns & Hansbrough, 2012). However, the failure of our replication should prompt future studies should look to replicate a wider range of designs relating to ROL to determine more conclusively whether ROL still exists.

Strengths & Limitations

The design of our replication has several strengths, the most notable being that this replication utilized a registered research approach in which a proposal was submitted before data collection. To combat the problems of “significosis” and “arigorium” that are common in leadership research (Antonakis, 2017), our proposal was peer-reviewed and included a thorough description of the design, sampling approach, sample size, methods, and specific empirical tests (including syntax) that we would utilize. Thus, we faithfully followed our intended research plan that was established at the onset of this replication endeavor. Over four studies and almost 2,000 participants representing a diverse sample of age, gender, work experience, and tenure, we were unable to replicate the findings of a smaller, less diverse sample. We even adapted the overall conditions to incentivize a closer reading of the narrative and evaluation of the numbers, and still received the same results. We feel confident that in replicating our outcomes over these four studies that we cannot replicate Meindl & Ehrlich’s experiment.

Despite the strength of our registered research design, multiple studies, and strong participation, there are limitations to the current study. Our studies were limited by the use of a fictional organizational vignette, rather than a real organization that may be meaningful for participants. Meindl and Ehrlich (1987) themselves noted this limitation and recommended: “future efforts might benefit from examining such processes in real, ongoing events and

evaluations, using individuals and groups with both vested interests in veridical views of organization function and some expertise in judging relevant cause-effect linkages in organized systems” (p. 106-107).

In addition, despite our efforts to recruit a large sample of MBA students using a similar procedure to Meindl and Ehrlich, we fell short of our desired sample size. However, our achieved sample size (N=292) was still well beyond that of Meindl and Ehrlich (N=111). Our low response rate may be in part due to a documented increase in survey non-response rates over the past several decades (Laguilles, Williams, & Saunders, 2011; Loosveldt & Storms, 2008; Pike, 2008; Tourangeau, 2004). Additionally, we decided in Study 1 not to provide any incentives for the participants (to match the approach of Meindl & Ehrlich, 1987), which may have also contributed to the low response rate. For example, Bentley and Thacker found that monetary payment had a significant effect on willingness to participate in research, with “lower levels of payment leading to lower willingness ratings” (Bentley & Thacker, 2004, p. 296). Future research might consider using incentives for MBA students, such as gift card draws to increase the response rate.

Next, our use of a crowdsourcing platform for participant recruitment in Studies 2-4 may have been limiting. The merits of using online data collection platforms such as Respondi, MTURK, or Qualtrics have been debated. However, accumulating research suggests that participants are more attentive to instructions than undergraduate student samples (Hauser & Schwarz, 2016). Others have also noted these platforms “can offer access to reliable and high-quality data from participants whose demographics resemble the general labor force more than student samples (e.g., older, more diverse)” (Roulin, 2015, p. 195), which helped us achieve the goals of a conceptual replication.

Finally, our use of behavioral intentions could be a limiting factor when examining individuals' true responses to ROL. In Study 4, we included additional outcome variables on behavioral intentions of the participants' willingness to support or invest in these companies. However, these are only intentions, rather than enacted behaviors. Because behavioral intentions are not costly to the participant involving real-world tradeoffs, they may be more prone to demand effects or social desirability bias (Lonati, Quiroga, Zehnder, & Antonakis, 2018). We recommend future research addresses more costly outcomes such as financial incentives or other types of extrinsic or intrinsic motivators.

Replication and Retesting of Leadership Models

Although replication is necessary for advancing leadership science, we found few examples in the extant leadership literature to help guide our ROL replication effort. Thus, we drew heavily from commentaries published in the fields of general psychology and experimental social psychology (Brandt et al., 2014; Crandall & Sherman, 2016; Erdfelder & Ulrich, 2018; Hüffmeier, Mazei, & Schultze, 2016; Schmidt, 2009) and even sourced articles published in languages other than English. We encourage future leadership replications to follow the steps we laid out in our replication approach section and to explore both types of replications: close and conceptual (Brandt et al., 2014). Our use of both close and conceptual replication allowed us to not only test whether or not the results of Meindl and Ehrlich (1987) hold but also to a) estimate the effect sizes expected in further replications, b) include additional conditions to address criticisms of the original work, and c) understand the generalizability of ROL to broader contexts.

CONCLUSION

By conducting a replication of Meindl and Ehrlich's (1987) study, we intended to rekindle a dialogue regarding the Romance of Leadership. ROL may continue to contribute to our collective understanding of how events are explained and why so much power is attributed to leaders; however, we consistently found that these attributions did not extend to evaluations of company performance. Perhaps we may look at this as "good news" suggesting a more objective approach to evaluating company performance with less of a leadership attribution bias. Interestingly, the strongest effects were with the MBA sample, which might serve as a caution against a primary focus of developing individual-level leadership behaviors in business schools (Collinson & Tourish, 2015; Petriglieri & Petriglieri, 2015) and an obsession with charismatic and transformational leaders (Bligh et al., 2011; van Knippenberg & Sitkin, 2013). In conclusion, our failure to replicate Meindl & Ehrlich's (1987) study should spark future research to examine in more detail the conditions under which individuals do and do not romanticize leaders and leadership and outcomes of such attributions.

REFERENCES

- Antonakis, J. (2017). On doing better science: From thrill of discovery to policy implications. *The Leadership Quarterly*, 28(2), 5-21. doi: 10.1016/j.leaqua.2017.01.006
- Bentley, J. P., & Thacker P.G. (2004). The influence of risk and monetary payment on the research participation decision making process. *Journal of Medical Ethics*, 30, 293-298.
- Bligh, M. C., Kohles, J. C., & Meindl, J. R. (2004). Charting the language of leadership: A methodological investigation of President Bush and the crisis of 9/11. *Journal of Applied Psychology*, 89, 562-574. doi:10.1037/0021-9010.89.3.562
- Bligh, M. C., Kohles, J. C., Pearce, C. L., Justin, J. E., & Stovall, J. F. (2007). When the romance is over: Follower perspectives of aversive leadership. *Applied Psychology: An International Review*, 56, 528-557. doi:10.1111/j.1464-0597.2007.00303.x
- Bligh, M. C., Kohles, J. C., & Pillai, R. (2011). Romancing leadership: Past, present, and future. *The Leadership Quarterly*, 22, 1058-1077. <http://dx.doi.org/10.1016/j.leaqua.2011.09.003>
- Bligh, M. C., & Schyns, B. (2007). Leading question: The romance lives on: Contemporary issues surrounding the romance of leadership. *Leadership*, 3, 343-360. <https://doi.org/10.1177/1742715007079316>
- Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J. & Giner-Sorolla, R., Grange, J. A., Perugini, M., Spies, J. R., & van't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, 50, 217-224. doi:10.1016/j.jesp.2013.10.005
- 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, 1640-1642. doi:10.1177/0956797610384744

- Carsten, M. K., & Uhl-Bien, M. (2013). Ethical followership: An Examination of Followership Beliefs and Crimes of Obedience. *Journal of Leadership & Organizational Studies*, 20, 49-61. doi:10.1177/0956797610384744
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Collinson, D., Jones, O. S., & Grint, K. (2018). 'No more heroes': Critical perspectives on leadership romanticism. *Organization Studies*, 39, 1625-1647. doi:10.1177/0170840617727784
- Collinson, D., & Tourish, D. (2015). Teaching leadership critically: New directions for leadership pedagogy. *Academy of Management Learning & Education*, 14, 576-594. <http://dx.doi.org/10.5465/amle.2014.0079>
- Crandall, C., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology*, 66, 93-99. doi:10.1016/j.jesp.2015.10.002
- Day, D. V., & Harrison, M. M. (2007). A multilevel, identity-based approach to leadership development, *Human Resource Management Review*, 17, 360-373.
- Eagly, A. H., & Johnson, B. T. (1990). Gender and leadership style: A meta-analysis. *Psychological Bulletin*, 108, 233-256. <http://dx.doi.org/10.1037/0033-2909.108.2.233>.
- Eagly, A. H., Karau, S. J., & Makhijani, M. G. (1995). Gender and the effectiveness of leaders: a meta-analysis. *Psychological Bulletin*, 117, 125-145. <http://dx.doi.org/10.1037/0033-2909.117.1.125>
- Erdfelder, E., & Ulrich, R. (2018). Zur Methodologie von Replikationsstudien. (On the methodology of replication studies.) *Psychologische Rundschau*, 69, 3-21.

- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Felfe, J., & Petersen, L. E. (2007). Romance of leadership and management decision making. *European Journal of Work and Organizational Psychology*, *16*, 1-24. <https://doi.org/10.1080/13594320600873076>
- Felfe, J., & Schyns, B. (2014). Romance of leadership and motivation to lead. *Journal of Managerial Psychology*, *29*, 850-865. doi:10.1108/JMP-03-2012-0076
- Fletcher, J. K. (2004). The paradox of postheroic leadership: An essay on gender, power, and transformational change. *The Leadership Quarterly*, *15*, 647-661. doi:10.1016/j.leaqua.2004.07.004
- Gibson, D. E., & Schroeder, S. J. (2003). Who ought to be blamed? The effect of organizational roles on blame and credit attributions. *International Journal of Conflict Management*, *14*, 95-117. <https://doi.org/10.1108/eb022893>
- Gray, J. H., & Densten, I. L. (2007). How leaders woo followers in the romance of leadership. *Applied Psychology*, *56*, 558-581. <http://dx.doi.org/10.1111/j.1464-0597.2007.00304.x>
- Gruber, D. A., Smerek, R. E., Thomas-Hunt, M. C., & James, E. H. (2015). The real-time power of Twitter: Crisis management and leadership in an age of social media. *Business Horizons*, *58*, 163-172. DOI: 10.1016/j.bushor.2014.10.006.
- Harms, P. D., Wood, D., Landay, K., Lester, P. B., & Lester, G. V. (2018). Autocratic leaders and authoritarian followers revisited: A review and agenda for the future. *The Leadership Quarterly*, *29*, 105-122. doi:10.1016/j.leaqua.2017.12.007

- Haslam, S. A., Platow, M. J., Turner, J. C., Reynolds, K. J., McGarty, C., Oakes, P. J., ... & Veenstra, K. (2001). Social identity and the romance of leadership: The importance of being seen to be 'doing it for us.' *Group Processes & Intergroup Relations*, *4*, 191-205. <https://doi.org/10.1177/1368430201004003002>
- Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*, 400-407.
- Heilman, M. E., & Haynes, M. C. (2005). No credit where credit is due: attributional rationalization of women's success in male-female teams. *Journal of Applied Psychology*, *90*, 905. <https://doi.org/10.1037/0021-9010.90.5.905>
- Henry, P. J. (2008). Student sampling as a theoretical problem. *Psychological Inquiry*, *19*, 114-126. <https://doi.org/10.1080/10478400802049951>
- Herrmann, D., & Felfe, J. (2009). Romance of Leadership und die Qualität von Managemententscheidungen (Romance of Leadership and the quality of management decision-making). *Zeitschrift für Arbeits-und Organisationspsychologie A&O*, *53*, 163-176. <https://doi.org/10.1026/0932-4089.53.4.163>
- Hino, K., & Aoki, H. (2013). Romance of leadership and evaluation of organizational failure. *Leadership & Organization Development Journal*, *34*, 365-377. doi:10.1108/LODJ-08-2011-0079
- Hüffmeier, J., Mazei, J., & Schultze, T. (2016). Reconceptualizing replication as a sequence of different studies: A replication typology. *Journal of Experimental Social Psychology*, *66*, 81-92. doi:10.1016/j.jesp.2015.09.009

- Keppel, G. (1991). *Design and analysis: A Researcher's Handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Kim, Y. (2014). Strategic communication of corporate social responsibility (CSR): Effects of stated motives and corporate reputation on stakeholder responses. *Public Relations Review, 40*, 838-840.
- Kulich, C., Ryan, M. K., & Haslam, S. A. (2007). Where is the romance for women leaders? The effects of gender on leadership attributions and performance- based pay. *Applied Psychology: An International Review, 56*, 582-601. <https://doi.org/10.1111/j.1464-0597.2007.00305.x>
- Laguilles, J. S., Williams, E. A., & Saunders, D. B. (2011). Can lottery incentives boost web survey response rates? Findings from four experiments. *Research in Higher Education, 52*, 537-553.
- Liu, H., Cutcher, L., & Grant, D. (2017). Authentic leadership in context: An analysis of banking CEO narratives during the global financial crisis. *Human Relations, 70*, 694-724. doi:10.1177/0018726716672920
- Lonati, S., Quiroga, B. F., Zehnder, C., & Antonakis, J. (2018). On doing relevant and rigorous experiments: Review and recommendations. *Journal of Operations Management, 64*, 19-40.
- Loosveldt, G., & Storms, V. (2008). Measuring public opinions about surveys. *International Journal of Public Opinion Research, 20*, 74-89
- Lorenz-Spreen, P., Mønsted, B. M., Hövel, P., & Lehmann, S. (2019). Accelerating dynamics of collective attention. *Nature Communications, 10*, 1-9.

- Meindl, J. R. (1990). On leadership: An alternative to the conventional wisdom. *Research in Organizational Behavior*, 12, 159-203.
- Meindl, J. R. (1995). The romance of leadership as a follower-centric theory: A social constructionist approach. *The Leadership Quarterly*, 6, 329-341.
[https://doi.org/10.1016/1048-9843\(95\)90012-8](https://doi.org/10.1016/1048-9843(95)90012-8)
- Meindl, J. R., Ehrlich, S. B., & Dukerich, J. M. (1985). The romance of leadership. *Administrative Science Quarterly*, 30, 78-102. <http://dx.doi.org/10.2307/2392813>
- Meindl, J. R., & Ehrlich, S. B. (1987). The romance of leadership and the evaluation of organizational performance. *Academy of Management Journal*, 30, 91-109.
<http://dx.doi.org/10.2307/255897>
- Meindl, J. R., & Ehrlich, S. B. (1988). Developing a “romance of leadership” scale. *Proceedings of the Eastern Academy of Management*, 133–135.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17, 776-783.
- Park, S. H., & Westphal, J. D. (2013). Social discrimination in the corporate elite: How status affects the propensity for minority CEOs to receive blame for low firm performance. *Administrative Science Quarterly*, 58, 542-586. doi:10.1177/0001839213509364
- Passarelli, A. M., Boyatzis, R. E., & Wei, H. (2018). Assessing leader development: Lessons from a historical review of MBA outcomes. *Journal of Management Education*, 42, 55-79.

- Peck, J. A., & Hogue, M. (2018). Acting with the best of intentions... or not: A typology and model of impression management in leadership. *The Leadership Quarterly*, 29, 123-134. doi:10.1016/j.leaqua.2017.10.001
- Petriglieri, G., & Petriglieri, J. L. (2015). Can business schools humanize leadership?. *Academy of Management Learning & Education*, 14, 625-647. <http://dx.doi.org/10.5465/amle.2014.0201>
- Pew Research Center. (2015). *Women and Leadership*. Retrieved from <http://www.pewsocialtrends.org/2015/01/14/chapter-1-women-in-leadership/>
- Pike, G. R. (2008). Using weighting adjustments to compensate for survey non-response. *Research in Higher Education*, 49, 153-171
- Pillai, R., & Meindl, J. R. (1998). Context and charisma: A “meso” level examination of the relationship of organic structure, collectivism, and crisis to charismatic leadership. *Journal of Management*, 24, 643-671.
- Randle, M., Eckersley, R., & Miller, L. (2017). Societal and personal concerns, their associations with stress, and the implications for progress and the future. *Futures*, 93, 68-79. doi:10.1016/j.futures.2017.07.004
- Roulin, N. (2015). Don't throw the baby out with the bathwater: Comparing data quality of crowdsourcing, online panels, and student samples. *Industrial and Organizational Psychology*, 8, 190-196.
- Rousseau, D. M. (2006). Is there such a thing as “evidence-based management”?. *Academy of Management Review*, 31, 256-269.
- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13, 90-100. doi:10.1037/a0015108

- Schyns, B., Felfe, J., & Blank, H. (2007). Is charisma hyper-romanticism? Empirical evidence from new data and a meta-analysis. *Applied Psychology, 56*, 505-527. DOI 10.1111/j.1464-0597.2007.00302.x
- Schyns, B., & Hansbrough, T. (2012). The romance of leadership scale and causal attributions. *Journal of Applied Social Psychology, 42*, 1870-1886. <http://dx.doi.org/10.1111/j.1559-1816.2012.00922.x>
- Seltzer, M. (2018). Measuring bias against female political leadership. *Politics & Gender, 1-27*.
- Shamir, B. (1992). Attribution of influence and charisma to the leader: The romance of leadership revisited. *Journal of Applied Social Psychology, 22*, 386-407. <https://doi.org/10.1111/j.1559-1816.1992.tb01546.x>
- Shamir, B., Pillai, R., Bligh, M. C., & Uhl-Bien, M. (2007). *Follower-centered perspectives on leadership: A tribute to the memory of James R. Meindl*. Greenwich: IAP
- Sinha, P. N., Inkson, K., & Barker, J. R. (2012). Committed to a failing strategy: Celebrity CEO, intermediaries, media and stakeholders in a co-created drama. *Organization Studies, 33*(2), 223-245. <https://doi.org/10.1177/0170840611430591>
- Streb, M. J., Burrell, B., Frederick, B., & Genovese, M. A. (2008). Social desirability effects and support for a female American president. *Public Opinion Quarterly, 72*, 76-89.
- Tourangeau, R. (2004). Survey research and societal change. *Annual Review of Psychology, 55*, 775-80.
- Ulhøi, J. P., & Müller, S. (2014). Mapping the landscape of shared leadership: A review and synthesis. *International Journal of Leadership Studies, 8*, 66-87.

U.S. Department of Labor, Bureau of Labor Statistics. (2017). Table 3: Employment Status of the Civilian Noninstitutional Population by Age, Sex, and Race, *Current Population Survey, Household Data Annual Averages 2016*.

van Knippenberg, D., & Sitkin, S. B. (2013). A critical assessment of charismatic—transformational leadership research: Back to the drawing board? *Academy of Management Annals*, 7, 1-60. <https://doi.org/10.1080/19416520.2013.759433>

Yukl, G. (2002). *Leadership in Organizations* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.

Table 1

Intercorrelations, Means and Standard Deviations in Study 1

	Mean	S.D.	1	2	3	4	5	6	7	8	9
1. Age	31.35	7.90									
2. Work experience	6.07	6.56	.71**								
3. Government condition	.26	.44	-.11	-.14*							
4. Market condition	.24	.43	-.01	-.02	-.33**						
5. Employee condition	.25	.44	.06	.03	-.34**	-.32**					
6. Leadership condition	.25	.44	.06	.13*	-.34**	-.33**	-.34*				
7. Profitability	7.46	1.57	-.13*	-.05	.06	-.08	-.04	.06			
8. Risk	5.29	2.04	.02	-.02	-.14*	-.02	.12*	.05	-.01		
9. Overall performance	6.38	1.28	-.06	-.05	-.07	-.07	.06	.08	.61**	.79**	
10. Semantic differential scale	7.10	1.35	-.04	-.08	-.17*	-.00	.08	.09	.35**	.36**	.50**

Note. ** $p < 0.01$ level (2-tailed), * $p < 0.05$ level (2-tailed). Experimental conditions are dummy coded (e.g. 1=yes, 0=no). Consistent with Meindl & Ehrlich (1987), risk was recoded such that higher scores indicated lower risk and overall performance was combination of high profit and low risk.

Table 2

Mean Evaluations and Standard Deviations (S.D.) for Each Explanation Condition in Study 1

	Explanation Condition								Statistics ^b		
	Leadership		Employee		Market		Government		Univariate F ^c	p	η^2
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.			
Profitability	7.64	1.51	7.34	1.86	7.23	1.70	7.63	1.16	1.20 (1.22)	.31 (.30)	.01
Risk ^a	5.46	2.19	5.70	2.00	5.21	1.81	4.81	2.08	2.57 (2.86)	.05 (.04)	.03
Overall performance	6.55	1.33	6.52	1.34	6.22	1.52	6.22	1.30	1.44 (1.88)	.23 (.13)	.01
Semantic differential scale	7.30	1.35	7.29	1.45	7.10	1.27	6.73	1.37	3.05 (2.73)	.03 (.04)	.03

Note. Number of participants in each condition were: leadership, N= 74; employees, N=73; market, N = 70; government, N =75. Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes.

^aConsistent with Meindl & Ehrlich (1987), risk was recoded such that higher scores indicated lower risk and overall performance was combination of high profit and low risk.^b Inclusion of covariates age and work experience denoted in parentheses. Degrees of freedom for MANOVA = 3, 287 and Degrees of freedom for MANCOVA = 3, 269.

Table 3

Intercorrelations, Means and Standard Deviations in Study 2

	Mean	S.D.	1	2	3	4	5	6	7	8	9
1. Age	45.45	10.75									
2. Work experience	14.36	10.24	.39**								
3. Government condition	.25	.43	-.09*	.02							
4. Market condition	.25	.43	.03	.03	-.34**						
5. Employee condition	.25	.43	-.01	-.04	-.33**	-.33**					
6. Leadership condition	.25	.43	.07	-.00	-.34**	-.34**	-.33**				
7. Profitability	7.46	1.66	.07	.03	-.06	.03	.02	.01			
8. Risk	6.14	2.20	-.05	-.00	-.02	-.02	-.02	.05	.03		
9. Overall performance	6.80	1.40	.01	.02	-.05	.00	-.00	.05	.62**	.81**	
10. Semantic differential scale	7.21	1.61	.09	.01	-.09	-.03	.04	.08	.56**	.31**	.57**

Note. ** $p < 0.01$ level (2-tailed). Experimental conditions are dummy coded (e.g. 1=yes, 0=no). Consistent with Meindl & Ehrlich (1987), risk was recoded such that higher scores indicated lower risk and overall performance was combination of high profit and low risk.

Table 4

Mean Evaluations and Standard Deviations (S.D.) for Each Explanation Condition in Study 2

	Explanation Condition								Statistics ^b		
	Leadership		Employee		Market		Government		Univariate F ^c	p	η^2
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.			
Profitability	7.50	1.68	7.52	1.67	7.54	1.59	7.27	1.72	.63 (.53)	.60 (.66)	.00
Risk ^a	6.33	2.27	6.07	2.22	6.06	2.28	6.08	2.05	.38 (.44)	.77 (.72)	.00
Overall performance	6.91	1.41	6.80	1.37	6.80	1.49	6.68	1.34	.53 (.54)	.66 (.66)	.00
Semantic differential scale	7.43	1.55	7.52	1.67	7.54	1.59	7.27	1.72	1.74 (1.60)	.16 (.19)	.01

Note. Number of participants in each condition were: leadership, N= 113; employees, N=110; market, N = 112; government, N =113. Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes.

^a *Consistent with Meindl & Ehrlich (1987), risk was recoded such that higher scores indicated lower risk and overall performance was combination of high profit and low risk.*

^b *Inclusion of covariates age and work experience denoted in parentheses. Degrees of freedom in MANOVA= 3, 444 and MANCOVA = 3, 442.*

Table 5

Comparison of Meindl and Ehrlich (1987) Results and Effect Sizes with Close Replication Results

	Meindl & Ehrlich (1987) MBA Students N=110		Study 1 MBA Students N=292		Study 2 Working Adults N=448		Study 1 & 2 Combined N=740	
	F-Test	η^2	F-Test	η^2	F-Test	η^2	F-Test	η^2
Profitability	8.0, p <.01	0.09	1.20, p =.31	.01	.63, p =.60	.004	.27, p =.84	.001
Risk	5.71, p <.10	0.06	2.57, p =.05	.03	.38, p =.77	.003	1.43, p =.24	.006
Overall performance	6.86, p <.001	0.12	1.44, p =.23	.01	.53, p =.66	.004	1.46, p =.23	.006
Semantic differential	7.69, p <.05	0.07	3.05, p =.03	.03	1.74, p =.16	.012	4.18, p =.01	.017

Note. Replication Hypothesis 1: *Individuals presented with leadership-attributed causal accounts will evaluate firm performance more favorably in terms of a) profitability, b) risk, c) overall performance, and d) positivity than when the same performance outcomes are attributed to non-leadership factors.*

Combined Study 1 & 2 controlling for study. Degrees of freedom = 3, 735

Table 6

Intercorrelations, Means and Standard Deviations in Study 3

	Mean	S.D.	1	2	3	4	5	6	7	8	9	10	11
1. Age	41.35	11.85											
2. Tenure	9.43	8.68	.57**										
3. Market Failure Condition	.17	.38	-.01	-.06									
4. Market Success Condition	.16	.37	.01	-.02	-.20**								
5. Female Leader Failure Condition	.16	.37	-.04	-.02	-.20**	-.19**							
6. Male Leader Failure Condition	.17	.38	-.02	-.02	-.21**	-.20**	-.20**						
7. Female Leader Success Condition	.17	.38	.01	.04	-.21**	-.20**	-.20**	-.21**					
8. Male Leader Success Condition	.16	.37	.04	.08	-.20**	-.19**	-.20**	-.20**	-.20**				
9. Profitability	5.75	2.79	.09	.09	-.26**	.30**	-.28**	-.28**	.26**	.27**			
10. Risk	4.52	2.57	.02	.07	-.12**	.11**	-.18**	-.15**	.15**	.19**	.22**		
11. Overall performance	5.14	2.09	.08	.10*	-.25**	.27**	-.30**	-.28**	.27**	.30**	.80**	.76**	
12. Semantic differential scale	5.64	2.52	.03	.10*	-.26**	.27**	-.35**	-.28**	.33**	.29**	.75**	.40**	.75**

Note. ** $p < 0.01$ level (2-tailed), * $p < 0.05$ level (2-tailed); Experimental conditions are dummy coded (e.g. 1=yes, 0=no). Risk was recoded such that higher scores indicated lower risk. Consistent with Meindl & Ehrlich (1987), risk was recoded such that higher scores indicated lower risk and overall performance was combination of high profit and low risk.

Table 7a

MAN(C)OVA Results for Study 3

	Profitability						Risk ^a					
	MANOVA			MANCOVA			MANOVA			MANCOVA		
	F	p	η^2	F	p	η^2	F	p	η^2	F	p	η^2
Intercept	4024.61	.00	.87	224.93	.00	.28	2044.78	.00	.78	146.34	.00	.20
Age				2.06	.15	.00				.29	.59	.00
Tenure				.18	.67	.00				1.44	.23	.00
Failure condition ^b	361.86	.00	.38	356.82	.00	.38	73.79	.00	.11	72.01	.00	.11
Causal account ^c	.59	.55	.00	.60	.55	.00	.60	.55	.00	.58	.56	.00
Failure x causal account ^d	.02	.98	.00	.02	.98	.00	.92	.40	.00	.90	.41	.00

Note. Number of participants in each condition were: male leadership success, $N = 93$; male leadership failure, $N = 102$; female leadership success, $N = 103$; female leadership failure, $N = 94$; market success, $N = 97$; market failure, $N = 100$.

^a Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes. Consistent with Meindl & Ehrlich (1987), risk was recoded such that higher scores indicated lower risk and overall performance was combination of high profit and low risk.

^b Degrees of freedom in MANOVA = 1, 583 and MANCOVA = 1, 581.

^c Degrees of freedom in MANOVA = 2, 583 and MANCOVA = 2, 581.

^d Degrees of freedom in MANOVA = 2, 583 and MANCOVA = 2, 581.

Table 7a, *continued*.*MAN(C)OVA Results for Study 3*

	Overall Performance						Semantic Differential					
	MANOVA			MANCOVA			MANOVA			MANCOVA		
	F	p	η^2	F	p	η^2	F	p	η^2	F	p	η^2
Intercept	5716.31	.00	.91	357.88	.00	.38	5311.88	.00	.90	395.82	.00	.41
Age				.31	.58	.00				1.49	.22	.00
Tenure				1.36	.24	.00				4.15	.04	.01
Failure condition ^c	361.65	.00	.38	355.50	.00	.38	467.28	.00	.45	461.56	.00	.44
Causal account ^d	.86	.42	.00	.87	.42	.00	.53	.59	.00	.60	.55	.00
Failure x causal account ^e	.45	.64	.00	.41	.66	.00	2.30	.10	.01	2.33	.10	.01

Note. ^a Number of participants in each condition were: male leadership success, $N = 93$; male leadership failure, $N = 102$; female leadership success, $N = 103$; female leadership failure, $N = 94$; market success, $N = 97$; market failure, $N = 100$.

^a Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes. Consistent with Meindl & Ehrlich (1987), risk was recoded such that higher scores indicated lower risk and overall performance was combination of high profit and low risk.

^b Degrees of freedom in MANOVA = 1, 583 and MANCOVA = 1, 581.

^c Degrees of freedom in MANOVA = 2, 583 and MANCOVA = 2, 581.

^d Degrees of freedom in MANOVA = 2, 583 and MANCOVA = 2, 581.

Table 7b

Mean Evaluations and Standard Deviations (S.D.) for Each Explanation Condition in Study 3

	Explanation Condition Mean ^a											
	Male Leadership Success		Female Leadership Success		Market Success		Male Leadership Failure		Female Leadership Failure		Market Failure	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Profitability	7.52	1.67	7.34	1.90	7.62	1.78	4.03	2.55	3.94	2.60	4.14	2.49
Risk	5.62	2.20	5.35	2.47	5.18	2.19	3.71	2.42	3.45	2.68	3.84	2.54
Overall performance	6.57	1.42	6.34	1.42	6.40	1.36	3.87	1.84	3.69	1.90	3.99	1.86
Semantic differential scale	7.35	1.77	7.43	1.65	7.18	1.72	4.09	2.13	3.64	1.98	4.19	1.98

Note. ^a Number of participants in each condition were: male leadership success, $N = 93$; male leadership failure, $N = 102$; female leadership success, $N = 103$; female leadership failure, $N = 94$; market success, $N = 97$; market failure, $N = 100$.

Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes. Consistent with Meindl & Ehrlich (1987), risk was recoded such that higher scores indicated lower risk and overall performance was combination of high profit and low risk.

Table 8
Intercorrelations, Means and Standard Deviations in Study 4

	Mean	S.D.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Age	44.78	11.34															
2. Work Experience	18.21	11.41	.75**														
3. Male Leader Success Condition	.17	.38	.01	.03													
4. Male Leader Failure Condition	.15	.36	-.04	-.02	-.19**												
5. Female Leader Success Condition	.18	.38	.00	.03	-.21**	-.20**											
6. Female Leader Failure Condition	.17	.38	.02	-.04	-.21**	-.19**	-.21**										
7. Market Success Condition	.19	.39	.04	.02	-.22**	-.21**	-.23**	-.22**									
8. Market Failure Condition	.14	.35	-.04	-.02	-.18**	-.17**	-.19**	-.18**	-.20**								
9. Profitability	5.94	2.69	.07	.00	.22**	-.29**	.28**	-.33**	.25**	-.18**							
10. Risk	4.56	2.41	-.03	-.03	.13**	-.15**	.08*	-.12**	.16**	-.13**	.15**						
11. Overall performance	1.49	3.87	.02	-.01	.23**	-.29**	.24**	-.30**	.27**	-.20**	.79**	.73**					
12. Semantic differential scale	5.63	2.31	.05	.02	.25**	-.33**	.31**	-.33**	.27**	-.22**	.70**	.37**	.71**				
13. Support intentions	3.13	1.04	-.05	-.09*	.22**	-.24**	.21**	-.32**	.24**	-.15**	.69**	.27**	.64**	.71**			
14. Investment intentions	3.17	1.19	.00	-.01	.23**	-.28**	.29**	-.32**	.26**	-.22**	.72**	.28**	.67**	.74**	.86**		
15. Employment intentions	3.05	1.00	-.04	-.11*	.18**	-.24**	.22**	-.27**	.23**	-.16**	.66**	.22**	.59**	.67**	.89**	.82**	
16. Purchase intentions	3.11	.98	-.02	-.06	.18**	-.21**	.19**	-.24**	.19**	-.14**	.60**	.22**	.56**	.62**	.84**	.77**	.82**

Note. ** $p < 0.01$ level (2-tailed), * $p < 0.05$ level (2-tailed). Experimental conditions are dummy coded (e.g. 1=yes, 0=no). Risk was recoded such that higher scores indicated lower risk and overall performance was combination of high profit and low risk.

Table 9a

MAN(C)OVA Results for Each Explanation Condition in Study 4

	Profitability						Risk ^a					
	MANOVA			MANCOVA			MANOVA			MANCOVA		
	F	p	η^2	F	p	η^2	F	p	η^2	F	p	η^2
Intercept	4130.72	.00	.88	145.28	.00	.20	2222.40	.00	.79	131.24	.00	.19
Age				7.31	.01	.01				.62	.43	.00
Work experience				7.08	.01	.01				.02	.90	.00
Failure condition ^b	284.66	.00	.33	291.24	.00	.34	52.93	.00	.08	53.75	.00	.09
Causal account ^c	1.50	.23	.01	1.67	.19	.01	.16	.86	.00	.15	.86	.00
Failure x causal account ^d	2.50	.08	.01	3.31	.04	.01	.81	.45	.00	.61	.54	.00

Note. Number of participants in each condition were: male leadership success, $N = 101$; male leadership failure, $N = 89$; female leadership success, $N = 105$; female leadership failure, $N = 101$; market success, $N = 114$; market failure, $N = 84$. Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes.

^a Consistent with Meindl & Ehrlich (1987), risk was recoded such that higher scores indicated lower risk and overall performance was combination of high profit and low risk.

^b Degrees of freedom in MANOVA = 1, 583 and MANCOVA = 1, 581.

^c Degrees of freedom in MANOVA = 2, 583 and MANCOVA = 2, 581.

^d Degrees of freedom in MANOVA = 2, 583 and MANCOVA = 2, 581.

Table 9a, cont.

MAN(C)OVA Results for Each Explanation Condition in Study 4

	Overall Performance ^a						Semantic Differential					
	MANOVA			MANCOVA			MANOVA			MANCOVA		
	F	p	η^2	F	p	η^2	F	p	η^2	F	p	η^2
Intercept	6251.80	.00	.91	281.56	.00	.33	5718.72	.00	.91	227.74	.00	.28
Age				1.68	.20	.00				3.37	.07	.01
Work experience				3.76	.05	.01				3.14	.08	.01
Failure condition ^b	290.02	.00	.33	296.38	.00	.34	411.05	.00	.41	413.57	.00	.42
Causal account ^c	1.27	.28	.00	1.23	.29	.00	1.52	.22	.01	1.53	.22	.01
Failure x causal account ^d	.47	.63	.00	.60	.55	.00	2.09	.13	.01	2.64	.07	.01

Note. Number of participants in each condition were: male leadership success, $N = 101$; male leadership failure, $N = 89$; female leadership success, $N = 105$; female leadership failure, $N = 101$; market success, $N = 114$; market failure, $N = 84$. Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes.

^a Consistent with Meindl & Ehrlich (1987), risk was recoded such that higher scores indicated lower risk and overall performance was combination of high profit and low risk.

^b Degrees of freedom in MANOVA = 1, 583 and MANCOVA = 1, 581.

^c Degrees of freedom in MANOVA = 2, 583 and MANCOVA = 2, 581.

^d Degrees of freedom in MANOVA = 2, 583 and MANCOVA = 2, 581.

Table 9a, cont.

MAN(C)OVA Results for Each Explanation Condition in Study 4

	Support Intentions						Investment Intentions					
	MANOVA			MANCOVA			MANOVA			MANCOVA		
	F	p	η^2	F	p	η^2	F	p	η^2	F	p	η^2
Intercept	721.17	.00	.93	367.07	.00	.39	6272.59	.00	.91	302.01	.00	.34
Age				.64	.42	.00				.26	.61	.00
Work experience				8.32	.00	.01				1.85	.18	.00
Failure condition ^b	214.08	.00	.27	217.52	.00	.27	322.12	.00	.35	323.50	.00	.36
Causal account ^c	2.44	.09	.01	2.70	.07	.01	.55	.58	.00	.56	.57	.00
Failure x causal account ^d	1.59	.21	.01	1.97	.14	.01	1.23	.29	.00	1.61	.20	.01

Note. Number of participants in each condition were: male leadership success, $N = 101$; male leadership failure, $N = 89$; female leadership success, $N = 105$; female leadership failure, $N = 101$; market success, $N = 114$; market failure, $N = 84$. Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes.

^a Consistent with Meindl & Ehrlich (1987), risk was recoded such that higher scores indicated lower risk and overall performance was combination of high profit and low risk.

^b Degrees of freedom in MANOVA = 1, 583 and MANCOVA = 1, 581.

^c Degrees of freedom in MANOVA = 2, 583 and MANCOVA = 2, 581.

^d Degrees of freedom in MANOVA = 2, 583 and MANCOVA = 2, 581.

Table 9a, cont.
MAN(C)OVA Results for Each Explanation Condition in Study 4

	Employment Intentions						Purchase Intentions					
	MANOVA			MANCOVA			MANOVA			MANCOVA		
	F	p	η^2	F	p	η^2	F	p	η^2	F	p	η^2
Intercept	6997.87	.00	.92	330.32	.00	.36	7032.20	.00	.92	331.89	.00	.37
Age				3.02	.08	.01				1.14	.29	.00
Work experience				14.25	.00	.02				5.57	.02	.01
Failure condition ^b	182.83	.00	.24	190.21	.00	.25	131.74	.00	.18	131.73	.00	.19
Causal account ^c	1.03	.36	.00	1.17	.31	.00	.61	.54	.00	.72	.49	.00
Failure x causal account ^d	.66	.52	.00	1.04	.35	.00	.73	.48	.00	.93	.40	.00

Note. Number of participants in each condition were: male leadership success, $N = 101$; male leadership failure, $N = 89$; female leadership success, $N = 105$; female leadership failure, $N = 101$; market success, $N = 114$; market failure, $N = 84$. Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes.

^a Consistent with Meindl & Ehrlich (1987), risk was recoded such that higher scores indicated lower risk and overall performance was combination of high profit and low risk.

^b Degrees of freedom in MANOVA = 1, 583 and MANCOVA = 1, 581.

^c Degrees of freedom in MANOVA = 2, 583 and MANCOVA = 2, 581.

^d Degrees of freedom in MANOVA = 2, 583 and MANCOVA = 2, 581.

Table 9b

Mean Evaluations and Standard Deviations (S.D.) for Each Explanation Condition in Study 4

	Explanation Condition Mean ^a											
	Male Leadership Success		Female Leadership Success		Market Success		Male Leadership Failure		Female Leadership Failure		Market Failure	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Profitability	7.23	2.04	7.53	1.64	7.32	1.89	4.11	2.53	4.01	2.51	4.77	2.58
Risk	5.25	2.22	4.98	2.22	5.35	2.36	3.69	2.31	3.93	2.56	3.80	2.16
Overall performance	6.24	1.50	6.26	1.30	6.33	1.43	3.90	1.74	3.97	1.85	4.29	1.70
Semantic differential scale	6.88	1.61	7.19	1.46	6.92	1.81	3.81	1.83	3.94	1.97	4.38	1.92
Support intentions	3.64	.77	3.60	.70	3.64	.76	2.54	.96	2.40	1.02	2.75	1.10
Investment intentions	3.76	.83	3.90	.74	3.81	.82	2.38	1.07	2.32	1.14	2.53	1.12
Employment intentions	3.45	.74	3.52	.72	3.52	.72	2.47	.96	2.45	1.01	2.65	1.09
Purchase intentions	3.50	.80	3.50	.75	3.49	.76	2.61	1.00	2.58	1.04	2.77	.98

Note. Number of participants in each condition were: male leadership success, $N = 101$; male leadership failure, $N = 89$; female leadership success, $N = 105$; female leadership failure, $N = 101$; market success, $N = 114$; market failure, $N = 84$.

^a Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes. Consistent with Meindl & Ehrlich (1987), risk was recoded such that higher scores indicated lower risk and overall performance was combination of high profit and low risk.

APPENDIX A – Studies 1 and 2 Materials

Instructions: The purpose of this questionnaire is to examine the effectiveness of certain business and financial information in allowing people to form general impressions and evaluations of firms traded on the New York Stock Exchange. As such, you will be asked to evaluate a firm based on information from several business journals and investment references that have been condensed into the following fact sheet. Please review the information about the firm ACTRON carefully and provide your assessment of the firm's performance.

ACTRON Fact Sheet

Business Overview

ACTRON is a multinational pharmaceutical company. The Company develops prescription medicines and vaccines to improve human health. The Company's operations are organized into two divisions: Therapeutic Medicines and Preventive Treatments. The ACTRON Therapeutic Medicines division develops and commercializes prescription medicines in the areas of oncology, internal medicine, immunology, and rare diseases. The ACTRON Preventive Treatments division researches and develops vaccines to eradicate treatable diseases.

Key Operating Strengths

Condition 1 – Leadership account. ACTRON's performance over the last few years is largely attributable to the outstanding leadership of Ross M. Pearson-Chief Executive Officer and Chairman of the Board-and his cadre of senior executives who took over the reins of the firm in 2014. The positive effects of leadership can be seen in several key strategic decisions to complete the development and bring to market a number of new pharmaceuticals-particularly a line of dietary drugs-that have since proven to be highly successful.

Condition 2 - Employee account. ACTRON's performance over the last few years is largely attributable to the exceptional quality of the labor force it employs, particularly the fine new core of bench scientists-mostly biochemists and biochemical engineers-that came to the company and congealed into a productive group in 2014. The positive effects can be seen on a number of new and innovative product development efforts-particularly a line of dietary drugs-that have been brought to market and have since proven to be highly successful.

Condition 3 - Market account. ACTRON's performance in the last few years is largely attributable to external changes in market demand that became apparent in 2014. These external changes in consumer demand have favored Actron's particular mix of products, which is somewhat unusual in the industry. The positive effects can be seen, for example, in a number of its dietary drugs. These drugs-most of which had been developed and patented years ago-have suddenly within the last few years become highly successful.

Condition 4 - Government account. ACTRON's performance over the last few years is largely attributable to favorable changes in Federal drug regulatory policies that went into effect in 2014. The positive effects of these external changes on this firm's performance can be seen on a number of new pharmaceuticals-particularly a line of dietary drugs-that had been held up and restricted for some time. Deregulation in this specific area has fortuitously coincided with Actron's product development efforts-which have been atypical for the industry-allowing it to complete testing of new drugs and bring many of them to market. The new dietary drugs that have emerged from this situation have since proven to be highly successful.

Five-Year Summary Performance Indicators

	2013	2014	2015	2016	2017
Total Sales (in millions)	51,584	52,163	52,832	53,567	54,213
Profit Margins	15,917	16,465	16,973	17,469	18,219
Net Earnings	11,380	11,798	12,396	12,621	13,345
Earnings per Share	1.36	1.41	1.45	1.49	1.75
Stock Price	45.35	52.21	55.75	59.23	62.96

APPENDIX B – Studies 3 and 4 Materials

Instructions: The purpose of this questionnaire is to examine the effectiveness of certain business and financial information in allowing people to form general impressions and evaluations of firms traded on the New York Stock Exchange. As such, you will be asked to evaluate a firm based on information from several business journals and investment references that have been condensed into the following fact sheet. Please review the information about the firm ACTRON carefully and provide your assessment of the firm's performance.

ACTRON Fact Sheet

Business Overview

ACTRON is a multinational pharmaceutical company. The Company develops prescription medicines and vaccines to improve human health. The Company's operations are organized into two divisions: Therapeutic Medicines and Preventive Treatments. The ACTRON Therapeutic Medicines division develops and commercializes prescription medicines in the areas of oncology, internal medicine, immunology, and rare diseases. The ACTRON Preventive Treatments division researches and develops vaccines to eradicate treatable diseases.

Key Operating Strengths

Condition 1 – Male Leadership account X Success. ACTRON's performance over the last few years is largely attributable to the outstanding leadership of Ross M. Pearson-Chief Executive Officer and Chairman of the Board-and his cadre of senior executives who took over the reins of the firm in 2014. The positive effects of leadership can be seen in several key strategic decisions to complete the development and bring to market a number of new pharmaceuticals-particularly a line of dietary drugs-that have since proven to be highly successful.

Condition 2 – Male Leadership account X Failure. ACTRON's performance over the last few years is largely attributable to the poor leadership of Ross M. Pearson-Chief Executive Officer and Chairman of the Board-and his cadre of senior executives who took over the reins of the firm in 2014. The negative effects of leadership can be seen in several key strategic decisions to complete the development and bring to market a number of new pharmaceuticals-particularly a line of dietary drugs-that have since proven to be highly unsuccessful.

Condition 3 – Female Leadership account X Success. ACTRON's performance over the last few years is largely attributable to the outstanding leadership of Rose M. Pearson-Chief Executive Officer and Chairwoman of the Board-and her cadre of senior executives who took over the reins of the firm in 2014. The positive effects of leadership can be seen in several key strategic decisions to complete the development and bring to market a number of new pharmaceuticals-particularly a line of dietary drugs-that have since proven to be highly successful.

Condition 4 – Female Leadership account X Failure. ACTRON's performance over the last few years is largely attributable to the poor leadership of Rose M. Pearson-Chief Executive

Officer and Chairwoman of the Board-and her cadre of senior executives who took over the reins of the firm in 2014. The negative effects of leadership can be seen in several key strategic decisions to complete the development and bring to market a number of new pharmaceuticals-particularly a line of dietary drugs-that have since proven to be highly unsuccessful.

Condition 5 - Market account X Success. ACTRON's performance in the last few years is largely attributable to external changes in market demand that became apparent in 2014. These external changes in consumer demand have favored Actron's particular mix of products, which is somewhat unusual in the industry. The positive effects can be seen, for example, in a number of its dietary drugs. These drugs-most of which had been developed and patented years ago-have suddenly within the last few years become highly successful.

Condition 6 - Market account X Failure. ACTRON's performance in the last few years is largely attributable to external changes in market demand that became apparent in 2014. These external changes in consumer demand have not favored Actron's particular mix of products, which is somewhat unusual in the industry. The negative effects can be seen, for example, in a number of its dietary drugs. These drugs-most of which had been developed and patented years ago-have suddenly within the last few years become highly unsuccessful.

Five-Year Summary Performance Indicators – Success Conditions

	2013	2014	2015	2016	2017
Total Sales (in millions)	51,584	52,163	52,832	53,567	54,213
Profit Margins	15,917	16,465	16,973	17,469	18,219
Net Earnings	11,380	11,798	12,396	12,621	13,345
Earnings per Share	1.36	1.41	1.45	1.49	1.75
Stock Price	45.35	52.21	55.75	59.23	62.96

Five-Year Summary Performance Indicators – Failure Conditions

	2013	2014	2015	2016	2017
Total Sales (in millions)	51,584	50,163	48,832	47,567	45,213
Profit Margins	15,917	14,465	13,973	12,469	11,219
Net Earnings	11,380	11,098	10,396	10,121	9,345
Earnings per Share	1.36	1.31	1.25	1.19	1.05
Stock Price	45.35	42.21	39.75	35.23	32.96

Supplementary Material

Table 1

Mean Evaluations for Each Explanation Condition in Study 1 without Exclusion

Measures ^b	Explanation Condition Mean ^a								Statistics		
	Leadership		Employee		Market		Government		Univariate F ^C	p	η^2
	M	S.D.	M	S.D.	M	S.D.	M	S.D.			
Profitability	7.63	1.50	7.34	1.86	7.18	1.71	7.62	1.15	1.45 (1.31)	.23 (.27)	.02
Risk	5.45	2.17	5.70	2.00	5.19	1.83	4.83	2.13	2.49 (2.87)	.06 (.04)	.02
Overall performance	6.54	1.32	6.52	1.34	6.19	1.19	6.23	1.30	1.56 (2.01)	.20 (.11)	.02
Semantic differential scale	7.29	1.44	7.29	1.22	7.03	1.33	6.75	1.37	2.77 (2.60)	.04 (.05)	.03

Note. ^a Number of participants in each condition were: leadership, N= 75; employees, N=73; market, N = 72; government, N =77.

^b Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes. Risk was recoded such that higher scores indicate less risk.

^c Inclusion of covariates age and work experience denoted in parentheses. Degrees of freedom for MANOVA = 3, 293 and Degrees of freedom for MANCOVA = 3, 273.

Table 2

Mean Evaluations for Each Explanation Condition in Study 2 without Exclusion

Measures ^b	Explanation Condition Mean ^a								Statistics		
	Leadership		Employee		Market		Government		Univariate F ^c	p	η^2
	M	S.D.	M	S.D.	M	S.D.	M	S.D.			
Profitability	7.37	1.74	7.33	1.82	7.30	1.75	7.19	1.68	.27(.15)	.85 (.93)	.00
Risk	6.07	2.31	5.95	2.24	5.96	2.30	5.88	2.09	.17 (.16)	.92 (.92)	.00
Overall	6.72	1.40	6.64	1.38	6.63	1.47	6.53	1.32	.40 (.19)	.75 (.90)	.00
performance Semantic differential scale	7.26	1.60	7.05	1.69	6.93	1.66	6.80	1.60	1.88 (1.71)	.13 (.16)	.01

Note. ^a Number of participants in each condition were: leadership, N= 135; employees, N=141; market, N = 142; government, N =136.

^b Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes. Risk was recoded such that higher scores indicate less risk.

^c Inclusion of covariates age and work experience denoted in parentheses. Degrees of freedom in MANOVA= 3, 550 and MANCOVA = 3, 513.

Table 3a

MAN(C)OVA RESULTS for Study 3 without Exclusion

	Profitability						Risk ^a					
	MANOVA			MANCOVA			MANOVA			MANCOVA		
	F	p	η^2	F	p	η^2	F	p	η^2	F	p	η^2
Intercept	4289.65	.00	.87	252.89	.00	.28	2242.28	.00	.78	154.77	.00	.20
Age				1.54	.22	.00				1.10	.30	.00
Tenure				.18	.67	.00				1.41	.24	.00
Failure condition ^c	357.76	.00	.36	356.35	.00	.36	79.12	.00	.11	76.63	.00	.11
Causal account ^d	.17	.84	.00	.11	.90	.00	1.01	.37	.00	1.03	.36	.00
Failure x causal account ^e	.16	.85	.00	.18	.83	.00	1.47	.23	.01	1.37	.26	.00

Note. ^a Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes. Risk was recoded such that higher scores indicated lower risk.

^b Degrees of freedom in MANOVA= 1, 642, MANCOVA= 1, 639.

^c Degrees of freedom in MANOVA= 2, 642, MANCOVA= 1, 639.

^d Degrees of freedom in MANOVA= 2, 642, MANCOVA= 1, 639.

Table 3a, *continued*

MAN(C)OVA RESULTS for Study 3 without Exclusion

	Overall Performance ^b						Semantic Differential					
	MANOVA			MANCOVA			MANOVA			MANCOVA		
	F	p	η^2	F	p	η^2	F	p	η^2	F	p	η^2
Intercept	6321.15	.00	.91	400.12	.00	.39	5657.62	.00	.90	415.45	.00	.39
Age				.40	.53	.00				.89	.35	.00
Tenure				1.35	.25	.00				3.29	.07	.01
Failure condition ^c	378.21	.00	.37	373.01	.00	.37	439.01	.00	.41	433.20	.00	.40
Causal account ^d	1.01	.37	.00	.91	.40	.00	.32	.73	.00	.31	.73	.00
Failure x causal account ^e	.38	.68	.00	.34	.71	.00	2.44	.09	.01	2.52	.08	.01

Note. ^a Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes. Risk was recoded such that higher scores indicated lower risk.

^b Degrees of freedom in MANOVA= 1, 642, MANCOVA= 1, 639.

^c Degrees of freedom in MANOVA= 2, 642, MANCOVA= 1, 639.

^d Degrees of freedom in MANOVA= 2, 642, MANCOVA= 1, 639

Table 3b.

Mean Evaluations for Each Explanation Condition in Study 3 without Exclusion

Outcome ^b	Explanation Condition ^a											
	Male Leadership Success		Female Leadership Success		Market Success		Male Leadership Failure		Female Leadership Failure		Market Failure	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Profitability	7.42	1.78	7.32	1.93	7.53	1.77	4.17	2.63	4.06	2.60	4.06	2.50
Risk	5.62	2.19	5.25	2.50	5.12	2.18	3.66	2.38	3.37	2.64	3.89	2.54
Overall performance	6.52	1.42	6.28	1.41	6.33	1.38	3.92	1.81	3.72	1.89	3.98	1.83
Semantic differential scale	7.22	1.81	7.36	1.64	7.06	1.76	4.19	2.15	3.77	2.08	4.26	1.96

Note. ^a Number of participants in each condition were: male leadership success, $N = 104$; male leadership failure, $N = 110$; female leadership success, $N = 109$; female leadership failure, $N = 107$; market success, $N = 109$; market failure, $N = 109$.

^a Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes. Risk was recoded such that higher scores indicated lower risk.

Table 4a

MAN(C)OVA RESULTS for Study 4 without Exclusion

	Profitability						Risk ^a					
	MANOVA			MANCOVA			MANOVA			MANCOVA		
	F	p	η^2	F	p	η^2	F	p	η^2	F	p	η^2
Intercept	4367.17	.00	.87	15.47	.00	.20	2280.13	.00	.78	134.59	.00	.18
Age				8.92	.00	.01				1.10	.30	.00
Work												
Experience				8.37	.00	.01				.28	.60	.00
Failure												
condition ^c	309.18	.00	.33	316.63	.00	.34	47.55	.00	.07	46.89	.00	.07
Causal												
account ^d	1.21	.30	.00	1.33	.27	.00	.60	.55	.00	.55	.58	.00
Failure x												
causal												
account ^e	2.94	.05	.01	3.79	.02	.01	.83	.44	.00	.58	.56	.00

Note. Number of participants in each condition were: male leadership success, $N = 93$; male leadership failure, $N = 102$; female leadership success, $N = 103$; female leadership failure, $N = 94$; market success, $N = 97$; market failure, $N = 10$. Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes^a Risk was recoded such that higher scores indicated lower risk.

^b Consistent with Meindl & Ehrlich (1987), performance was combination of high profit and low risk

^c Degrees of freedom in MANOVA= 1, 627 and MANCOVA = 1, 614.

^d Degrees of freedom in MANOVA= 2, 627 and MANCOVA = 2, 614.

^e Degrees of freedom in MANOVA= 2, 627 and MANCOVA = 2, 614.

Table 4a, continued

MAN(C)OVA RESULTS for Study 4 without Exclusion

	Overall Performance ^b						Semantic Differential					
	MANOVA			MANCOVA			MANOVA			MANCOVA		
	F	p	η^2	F	p	η^2	F	p	η^2	F	p	η^2
Intercept	6584.34	.00	.91	292.37	.00	.32	6063.43	.00	.91	241.27	.00	.28
Age				1.66	.20	.00				3.56	.06	.01
Work experience				2.59	.11	.00				2.78	.10	.01
Failure condition ^c	299.40	.00	.32	302.49	.00	.33	413.37	.00	.40	416.14	.00	.40
Causal account ^d	1.67	.19	.01	1.66	.19	.01	2.07	.13	.01	2.03	.13	.01
Failure x causal account ^e	.65	.52	.00	.81	.45	.00	2.67	.07	.01	3.28	.04	.01

Note. Number of participants in each condition were: male leadership success, $N = 93$; male leadership failure, $N = 102$; female leadership success, $N = 103$; female leadership failure, $N = 94$; market success, $N = 97$; market failure, $N = 10$. Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes^a Risk was recoded such that higher scores indicated lower risk.

^b Consistent with Meindl & Ehrlich (1987), performance was combination of high profit and low risk

^c Degrees of freedom in MANOVA= 1, 627 and MANCOVA = 1, 614.

^d Degrees of freedom in MANOVA= 2, 627 and MANCOVA = 2, 614.

^e Degrees of freedom in MANOVA= 2, 627 and MANCOVA = 2, 614.

Table 4a, *continued*

MAN(C)OVA RESULTS for Study 4 without Exclusion

	Support Intentions						Investment Intentions					
	MANOVA			MANCOVA			MANOVA			MANCOVA		
	F	p	η^2	F	p	η^2	F	p	η^2	F	p	η^2
Intercept	7688.34	.00	.93	385.41	.00	.39	6652.38	.00	.91	311.54	.00	.34
Age				1.11	.29	.00				.65	.42	.00
Work experience				9.77	.00	.02				2.27	.13	.00
Failure condition ^c	225.75	.00	.27	229.36	.00	.27	326.94	.00	.34	328.47	.00	.35
Causal account ^d	3.07	.05	.01	3.27	.04	.01	.86	.43	.00	.84	.43	.00
Failure x causal account ^e	2.01	.14	.01	2.53	.08	.01	1.76	.17	.01	2.23	.11	.01

Note. Number of participants in each condition were: male leadership success, $N = 93$; male leadership failure, $N = 102$; female leadership success, $N = 103$; female leadership failure, $N = 94$; market success, $N = 97$; market failure, $N = 10$. Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes^a Risk was recoded such that higher scores indicated lower risk.

^b Consistent with Meindl & Ehrlich (1987), performance was combination of high profit and low risk

^c Degrees of freedom in MANOVA = 1, 627 and MANCOVA = 1, 614.

^d Degrees of freedom in MANOVA = 2, 627 and MANCOVA = 2, 614.

^e Degrees of freedom in MANOVA = 2, 627 and MANCOVA = 2, 614.

Table 4a, *continued**MAN(C)OVA RESULTS for Study 4 without Exclusion*

	Employment Intentions						Purchase Intentions					
	MANOVA			MANCOVA			MANOVA			MANCOVA		
	F	p	η^2	F	p	η^2	F	p	η^2	F	p	η^2
Intercept	7454.81	.00	.92	347.52	.00	.36	7407.23	.00	.92	341.94	.00	.36
Age				4.18	.04	.01				1.77	.18	.00
Work experience				17.20	.00	.03				6.26	.01	.01
Failure condition ^c	197.31	.00	.24	205.52	.00	.25	142.59	.00	.19	142.68	.00	.19
Causal account ^d	1.72	.18	.01	1.78	.17	.01	1.25	.29	.00	1.32	.27	.00
Failure x causal account ^e	.87	.42	.00	1.33	.27	.00	.74	.48	.00	.97	.38	.00

Note. Number of participants in each condition were: male leadership success, $N = 93$; male leadership failure, $N = 102$; female leadership success, $N = 103$; female leadership failure, $N = 94$; market success, $N = 97$; market failure, $N = 10$. Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes^a Risk was recoded such that higher scores indicated lower risk.

^b Consistent with Meindl & Ehrlich (1987), performance was combination of high profit and low risk

^c Degrees of freedom in MANOVA = 1, 627 and MANCOVA = 1, 614.

^d Degrees of freedom in MANOVA = 2, 627 and MANCOVA = 2, 614.

^e Degrees of freedom in MANOVA = 2, 627 and MANCOVA = 2, 614.

Table 4b

Mean Evaluations for Each Explanation Condition in Study 4 without Exclusion

Outcome ^b	Explanation Condition Mean ^a											
	Male Leadership Success		Female Leadership Success		Market Success		Male Leadership Failure		Female Leadership Failure		Market Failure	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Profitability	7.19	2.02	7.59	1.64	7.27	1.94	4.12	2.56	3.97	2.52	4.69	2.50
Risk	5.16	2.24	4.85	2.29	5.29	2.41	3.64	2.30	3.89	2.58	3.90	2.15
Overall	6.17	1.50	6.22	1.28	6.28	1.44	3.88	1.74	3.93	1.88	4.30	1.67
performance												
Semantic differential scale	6.75	1.64	7.15	1.47	6.88	1.82	3.86	1.84	3.90	1.99	4.42	1.82
Support intentions	3.62	.77	3.62	.72	3.66	.76	2.56	.95	2.39	1.03	2.77	1.08
Investment intentions	3.72	.85	3.88	.77	3.80	.82	2.40	1.07	2.31	1.14	2.56	1.09
Employment intentions	3.43	.74	3.55	.75	3.56	.74	2.49	.96	2.44	1.01	2.66	1.06
Purchase intentions	3.48	.80	3.50	.78	3.52	.77	2.59	1.00	2.57	1.05	2.79	.97

Note. Number of participants in each condition were: male leadership success, $N = 108$; male leadership failure, $N = 94$; female leadership success, $N = 111$; female leadership failure, $N = 102$; market success, $N = 124$; market failure, $N = 94$.

^a Response formats ranged from 0 to 10, with high scores indicating more favorable outcomes. Risk was recoded such that higher scores indicated lower risk. Consistent with Meindl & Ehrlich (1987), performance was combination of high profit and low risk

Table 5

Regression Analyses of Study 1

	Profitability						Risk ^a					
	Model 1			Model 2			Model 1			Model 2		
	B	SE	p	B	SE	p	B	SE	p	B	SE	p
Constant	8.46	.46	.00	8.63	.49	.00	4.78	.61	.00	5.11	.64	.00
Age	-.04	.02	.03	-.04	.02	.04	.02	.02	.34	.02	.02	.36
Work experience	.02	.02	.34	.02	.02	.36	-.03	.03	.36	-.03	.03	.24
Government condition ^c				-.02	.27	.93				-.80	.35	.02
Market condition ^d				-.44	.27	.11				-.37	.35	.30
Labor condition ^e				-.28	.27	.29				.14	.35	.70
R^2	.02			.03			.00			.04		
ΔR^2	.02			.01			.00			.03		
P-value of ΔR^2	.07			.30			.60			.04		

Note. $N = 279$.

^a Consistent with Meindl & Ehrlich (1987), risk was recoded such that high scores indicate lower risk.

^b Consistent with Meindl & Ehrlich (1987), performance was combination of high profit and lower risk.

^c Government condition dummy coded as 0 (all other conditions) and 1 (government condition).

^d Market condition dummy coded as 0 (all other conditions) and 1 (market condition).

^e Labor condition dummy coded as 0 (all other conditions) and 1 (labor condition).

Table 5, continued.

Regression Analyses of Study 1

	Overall Performance ^b						Semantic Differential					
	Model 1			Model 2			Model 1			Model 2		
	B	SE	p	B	SE	p	B	SE	p	B	SE	p
Constant	6.62	.38	.00	6.87	.40	.00	7.10	.39	.00	7.34	.41	.00
Age	-.01	.01	.59	-.01	.01	.58	.00	.01	.81	.00	.01	.83
Work experience	.00	.02	.88	-.01	.02	.71	-.02	.02	.28	-.02	.02	.17
Government condition ^c				-.41	.22	.07				-.56	.23	.02
Market condition ^d				-.41	.22	.07				-.21	.23	.35
Labor condition ^e				-.07	.22	.74				.01	.23	.98
R^2	.00			.02			.01			.04		
ΔR^2	.00			.02			.01			.03		
P-value of ΔR^2	.65			.13			.43			.04		

Note. $N = 279$.

^a Consistent with Meindl & Ehrlich (1987), risk was recoded such that high scores indicate lower risk.

^b Consistent with Meindl & Ehrlich (1987), performance was combination of high profit and lower risk.

^c Government condition dummy coded as 0 (all other conditions) and 1 (government condition).

^d Market condition dummy coded as 0 (all other conditions) and 1 (market condition).

^e Labor condition dummy coded as 0 (all other conditions) and 1 (labor condition).

Table 6

Regression Analyses of Study 2

	Profitability						Risk ^a					
	Model 1			Model 2			Model 1			Model 2		
	B	SE	p	B	SE	p	B	SE	p	B	SE	p
Constant	6.99	.35	.00	7.05	.38	.00	6.58	.46	.00	6.83	.51	.00
Age	.01	.01	.20	.01	.01	.26	-.01	.01	.31	-.01	.01	.27
Work experience	.00	.01	.89	.00	.01	.84	.00	.01	.70	.00	.01	.67
Government condition ^c				-.19	.22	.39				-.29	.30	.34
Market condition ^d				.06	.22	.81				-.28	.30	.35
Labor condition ^e				.04	.22	.86				-.27	.30	.36
R^2	.01			.01								
ΔR^2	.01			.00			.00			.01		
P-value of ΔR^2	.35			.66			.00			.00		

Note. $N = 448$.

^a Consistent with Meindl & Ehrlich (1987), risk was recoded such that high scores indicate low risk.

^b Consistent with Meindl & Ehrlich (1987), performance was combination of high profit and low risk.

^c Government condition dummy coded as 0 (all other conditions) and 1 (government condition).

^d Market condition dummy coded as 0 (all other conditions) and 1 (market condition).

^e Labor condition dummy coded as 0 (all other conditions) and 1 (labor condition).

Table 6, continued

Regression Analyses of Study 2

	Overall Performance ^b						Semantic Differential					
	Model 1			Model 2			Model 1			Model 2		
	B	SE	p	B	SE	p	B	SE	p	B	SE	p
Constant	6.78	.29	.00	6.94	.33	.00	6.63	.34	.00	6.90	.37	.00
Age	.00	.01	.96	.00	.01	.84	.01	.01	.13	.01	.01	.20
Work experience	.00	.01	.70	.00	.01	.65	.00	.01	.51	.00	.01	.42
Government condition ^c				-.24	.19	.20				-.42	.22	.05
Market condition ^d				-.11	.19	.56				-.29	.21	.18
Labor condition ^e				-.12	.19	.54				-.09	.22	.68
<i>R</i> ²												
ΔR^2	.00			.00			.01			.02		
P-value of ΔR^2	.00			.00			.01			.01		

Note. *N* = 448.

^a Consistent with Meindl & Ehrlich (1987), risk was recoded such that high scores indicate low risk.

^b Consistent with Meindl & Ehrlich (1987), performance was combination of high profit and low risk.

^c Government condition dummy coded as 0 (all other conditions) and 1 (government condition).

^d Market condition dummy coded as 0 (all other conditions) and 1 (market condition).

^e Labor condition dummy coded as 0 (all other conditions) and 1 (labor condition).

Table 7

Regression Analyses of Study 3

	Profitability									Risk ^a								
	Model 1			Model 2			Model 3			Model 1			Model 2			Model 3		
	B	SE	p	B	SE	p	B	SE	p	B	SE	p	B	SE	p	B	SE	p
Constant	4.98	.44	.00	6.88	.38	.00	6.89	.41	.00	4.49	.40	.00	5.57	.42	.00	5.68	.45	.00
Age	.02	.01	.21	.01	.01	.15	.01	.01	.15	-.01	.01	.66	-.01	.01	.60	-.01	.01	.59
Tenure	.00	.00	.25	.00	.00	.67	.00	.00	.67	.00	.00	.11	.00	.00	.23	.00	.00	.23
Failure Condition ^c				-3.44	.18	.00	-3.46	.32	.00				-1.70	.20	.00	-1.90	.35	.00
Female Leader Condition ^d				-.13	.22	.57	-.16	.32	.61				-.26	.25	.30	-.27	.35	.44
Market Condition ^e				.12	.22	.59	.13	.32	.69				-.13	.25	.60	-.42	.35	.23
Interaction of female leader condition x failure							.07	.45	.87							.00	.49	.99
Interaction of market condition x failure							-.01	.45	.98							.57	.49	.25
Total R ²	.01			.39			.39			.01			.12			.12		
ΔR ²	.01			.34			.00			.01			.11			.00		
P value of ΔR ²	.04			.00			.98			.23			.00			.41		

Note. $N = 589$. ^aConsistent with Meindl & Ehrlich (1987), risk was recoded such that high scores indicate low risk. ^bConsistent with Meindl & Ehrlich (1987), performance was combination of high profit and low risk. ^cFailure condition dummy coded as 0 (success) and 1 (failure). ^dFemale leader condition dummy coded as 0 (all other conditions) and 1 (female leader condition). ^eMarket condition dummy coded as 0 (all other conditions) and 1 (market condition).

Table 7, *continued**Regression Analyses of Study 3*

	Overall Performance ^b									Semantic Differential								
	Model 1			Model 2			Model 3			Model 1			Model 2			Model 3		
	B	SE	p	B	SE	p	B	SE	p	B	SE	p	B	SE	p	B	SE	p
Constant	4.73	.33	.00	6.22	.29	.00	6.28	.31	.00	5.66	.39	.00	7.56	.33	.00	7.52	.35	.00
Age	.01	.01	.58	.00	.01	.57	.00	.01	.58	-.01	.01	.43	-.01	.01	.23	-.01	.01	.22
Tenure	.00	.00	.08	.00	.00	.24	.00	.00	.24	.00	.00	.02	.00	.00	.04	.00	.00	.04
Failure Condition ^c				-2.57	.14	.00	-2.68	.24	.00				-3.33	.16	.00	-3.24	.27	.00
Female Leader Condition ^d				-.19	.17	.25	-.21	.24	.36				-.17	.19	.36	.09	.27	.74
Market Condition ^e				.00	.17	.98	-.15	.24	.54				.00	.19	.99	-.13	.27	.62
Interaction of female leader condition x failure							.04	.33	.91							-.54	.38	.16
Interaction of market condition x failure							.28	.33	.40							.26	.38	.49
Total R ²	.01			.39			.39						.44			.45		
ΔR^2	.01			.38			.00						.44			.00		
P value of ΔR^2	.04			.00			.66						.00			.10		

Note. $N = 589$. ^a Consistent with Meindl & Ehrlich (1987), risk was recoded such that high scores indicate low risk. ^b Consistent with Meindl & Ehrlich (1987), performance was combination of high profit and low risk. ^c Failure condition dummy coded as 0 (success) and 1 (failure). ^d Female leader condition dummy coded as 0 (all other conditions) and 1 (female leader condition). ^e Market condition dummy coded as 0 (all other conditions) and 1 (market condition).

Table 8
Regression Study 4

	Profitability									Risk ^a								
	Model 1			Model 2			Model 3			Model 1			Model 2			Model 3		
	B	SE	p	B	SE	p	B	SE	p	B	SE	p	B	SE	p	B	SE	p
Constant	5.08	.50	.00	6.43	.44	.00	6.36	.46	.00	5.00	.45	.00	5.69	.46	.00	5.78	.48	.00
Age	.03	.01	.07	.03	.01	.01	.03	.01	.01	-.01	.01	.43	-.01	.01	.48	-.01	.01	.43
Work Experience	-.02	.01	.19	-.03	.01	.01	-.03	.01	.01	.00	.01	.86	.00	.01	.84	.00	.01	.90
Failure Condition ^c				-3.10	.18	.00	-3.11	.32	.00				-1.41	.19	.00	-1.53	.34	.00
Female Leader Condition ^d				.08	.22	.73	.34	.31	.27				-.09	.24	.71	-.29	.32	.38
Market Condition ^e				.34	.22	.13	.09	.30	.77				.05	.24	.82	.08	.32	.81
Female leader condition x failure							-.53	.44	.23							.42	.47	.38
Market condition x failure							.59	.45	.19							-.06	.48	.91
Total R ²	.01			.34			.35			.00			.09			.09		
ΔR^2	.01			.34			.01			.00			.09			.09		
P value of ΔR^2	.18			.00			.04			.61			.00			.54		

Note. $N = 594$, ^a Consistent with Meindl & Ehrlich (1987), risk was recoded such that high scores indicate low risk. ^b Consistent with Meindl & Ehrlich (1987), performance was combination of high profit and low risk. ^c Failure condition dummy coded as 0 (success) and 1 (failure). ^d Female leader condition dummy coded as 0 (all other conditions) and 1 (female leader condition). ^e Market condition dummy coded as 0 (all other conditions) and 1 (market condition).

Table 8, *continued**Regression Study 4*

	Overall Performance ^b									Semantic Differential								
	Model 1			Model 2			Model 3			Model 1			Model 2			Model 3		
	B	SE	p	B	SE	p	B	SE	p	B	SE	p	B	SE	p	B	SE	p
Constant	5.04	.36	.00	6.06	.32	.00	6.07	.33	.00	5.13	.44	.00	6.41	.36	.00	6.40	.37	.00
Age	.01	.01	.43	.01	.01	.22	.01	.01	.20	.01	.01	.28	.02	.01	.10	.02	.01	.07
Work Experience	-.01	.01	.42	-.02	.01	.06	-.02	.01	.05	-.01	.01	.64	-.02	.01	.11	-.02	.01	.08
Failure Condition ^c				-2.25	.13	.00	-2.32	.23	.00				-3.00	.15	.00	-3.09	.26	.00
Female Leader Condition ^d				-.01	.16	.97	.03	.22	.91				.21	.18	.25	.34	.25	.17
Market Condition ^e				.20	.16	.22	.08	.22	.71				.28	.18	.13	.04	.24	.86
Interaction of female leader x failure							-.06	.32	.85							-.26	.36	.47
Interaction of market x failure							.27	.32	.41							.54	.37	.14
Total R ²	.00			.34			.35			.00			.42			.42		
ΔR^2	.00			.34			.01			.00			.42			.01		
P value of ΔR^2	.70			.00			.55			.49			.00			.07		

Note. $N = 594$.^a Consistent with Meindl & Ehrlich (1987), risk was recoded such that high scores indicate low risk. ^b Consistent with Meindl & Ehrlich (1987), performance was combination of high profit and low risk. ^c Failure condition dummy coded as 0 (success) and 1 (failure). ^d Female leader condition dummy coded as 0 (all other conditions) and 1 (female leader condition). ^e Market condition dummy coded as 0 (all other conditions) and 1 (market condition).

Table 8, *continued**Regression Study 4 (additional outcomes)*

	Support Intentions									Investment Intentions								
	Model 1			Model 2			Model 3			Model 1			Model 2			Model 3		
	B	SE	p	B	SE	p	B	SE	p	B	SE	p	B	SE	p	B	SE	p
Constant	3.22	.19	.00	3.73	.18	.00	3.72	.18	.00	3.17	.22	.00	3.82	.19	.00	3.77	.20	.00
Age	.00	.01	.75	.00	.00	.52	.00	.00	.42	.00	.01	.93	.00	.01	.72	.00	.01	.61
Work Experience	-.01	.01	.10	-.01	.00	.01	-.01	.00	.00	.00	.01	.80	-.01	.01	.22	-.01	.01	.17
Failure Condition ^c				-1.08	.07	.00	-1.09	.13	.00				-1.43	.08	.00	-1.38	.14	.00
Female Leader Condition ^d				-.10	.09	.27	-.03	.12	.83				.03	.10	.78	.14	.13	.29
Market Condition ^e				.09	.09	.32	.00	.12	.97				.09	.10	.36	.05	.13	.68
Interaction of female leader condition x failure							-.15	.18	.40							-.24	.19	.21
Interaction of market condition x failure							.20	.18	.27							.09	.20	.65
Total R ²	.01			.29			.29			.00			.36			.37		
ΔR ²	.01			.28			.00			.00			.36			.00		
P value of ΔR ²	.10			.00			.14			.96			.00			.20		

Note. N = 594.^a Consistent with Meindl & Ehrlich (1987), risk was recoded such that high scores indicate low risk. ^b Consistent with Meindl & Ehrlich (1987), performance was combination of high profit and low risk. ^c Failure condition dummy coded as 0 (success) and 1 (failure). ^d Female leader condition dummy coded as 0 (all other conditions) and 1 (female leader condition). ^e Market condition dummy coded as 0 (all other conditions) and 1 (market condition).

Table 8, continued.

Regression Study 4 (additional outcomes)

	Employment Intentions									Purchase Intentions								
	Model 1			Model 2			Model 3			Model 1			Model 2			Model 3		
	B	SE	p	B	SE	p	B	SE	p	B	SE	p	B	SE	p	B	SE	p
Constant	3.00	.19	.00	3.42	.17	.00	3.40	.18	.00	3.08	.18	.00	3.47	.18	.00	3.47	.19	.00
Age	.01	.01	.21	.01	.00	.11	.01	.00	.08	.00	.01	.48	.00	.00	.34	.01	.00	.29
Work Experience	-.01	.01	.01	-.02	.00	.00	-.02	.00	.00	-.01	.01	.13	-.01	.00	.02	-.01	.00	.02
Failure Condition ^c				-.99	.07	.00	-.99	.13	.00				-.85	.07	.00	-.87	.13	.00
Female Leader Condition ^d				.02	.09	.78	.09	.12	.46				-.03	.09	.74	.01	.12	.93
Market Condition ^e				.12	.09	.18	.07	.12	.56				.06	.09	.49	.00	.12	.97
Interaction of female leader condition x failure							-.13	.18	.45							-.08	.18	.64
Interaction of market condition x failure							.12	.18	.51							.16	.18	.39
Total R ²	.01			.26			.27			.01			.19			.20		
ΔR ²	.01			.25			.00			.01			.19			.00		
P value of ΔR ²	.02			.00			.35			.27			.00			.40		

Note. $N = 594$.^a Consistent with Meindl & Ehrlich (1987), risk was recoded such that high scores indicate low risk. ^b Consistent with Meindl & Ehrlich (1987), performance was combination of high profit and low risk. ^c Failure condition dummy coded as 0 (success) and 1 (failure). ^d Female leader condition dummy coded as 0 (all other conditions) and 1 (female leader condition). ^e Market condition dummy coded as 0 (all other conditions) and 1 (market condition).