Handelshøyskolen BI

GRA 19703 Master Thesis

Thesis Master of Science 100% - W

Predefinert informasjon

Startdato:	09-01-2023 09:00 CET	**Termin:**	202310
Sluttdato:	03-07-2023 12:00 CEST	**Vurderingsform:**	Norsk 6-trinns skala (A-F)
Eksamensform:	T		
Flowkode:	202310||11184||IN00||W||T		
Intern sensor:	(Anonymisert)		

Deltaker

Navn:	Aleksandra Groos Dyvik og Marie Louise Haraldsen

Informasjon fra deltaker

Tittel *:	Striking the Balance: Unveiling Interpersonal Consequences of Leader Overconfidence and the Potential Moderating Effect of Gender
Navn på veileder *:	Erik Løhre

Inneholder besvarelsen konfidensielt materiale?:	Nei	**Kan besvarelsen offentliggjøres?:**	Ja

---

Gruppe

Gruppenavn:	(Anonymisert)
Gruppenummer:	117
Andre medlemmer i gruppen:	

Master Thesis

BI Norwegian Business School

Striking the Balance: Unveiling Interpersonal Consequences of Leader Overconfidence and the Potential Moderating Effect of Gender

Exam code and name:

GRA 19703 – Master Thesis

Programme:

Master of Science in Leadership and Organizational

Psychology

Supervisor:
Erik Løhre

Hand-out date:
09.01.2023

Hand-in date:
03.07.2023

BI OSLO

Acknowledgment

We would like to take this opportunity to express our deepest gratitude to our supervisor, Erik Løhre, for his invaluable support, guidance, and constructive feedback throughout the entire master thesis. His expertise and dedication have been instrumental in shaping this research and enhancing its quality. We are truly grateful for the time and effort he invested in us.

Furthermore, we extend our sincere thanks to Veidekke and Gjensidige for providing us with the opportunity to collect the necessary data for our study. We greatly appreciate their collaboration and contribution to our research.

Moreover, we would also like to thank our families and friends for their unwavering support and understanding throughout this demanding academic journey. Their encouragement and belief in us have been a huge source of motivation.

Finally, we would like to acknowledge and thank each other for the remarkable cooperation, teamwork, and friendship that has defined this journey. Throughout countless hours of shared excitement, moments of frustration, engaging discussion, and endless amounts of coffee, we have consistently supported and motivated each other.

Abstract

Scholars hold contrasting views regarding the interpersonal consequences of overconfidence – some claim that overconfidence yields social advantages, whereas others contend that it results in adverse social outcomes. We examine the interpersonal effects of overconfidence in the context of high-level leadership and whether gender moderates the relationship between confidence level and interpersonal evaluations. A total of 454 participants, primarily recruited from two large and renowned companies in Norway, were randomly assigned to experimental conditions, wherein they watched an audio clip of either an (1) overconfident male CEO, (2) well-calibrated male CEO, (3) overconfident female CEO, or (4) well-calibrated female CEO. They were then asked to evaluate the CEO in terms of perceived competence and warmth before and after objective performance feedback was provided. The findings suggest that it is more beneficial for leaders to be well-calibrated than overconfident when performance is revealed to others. Contrary to conventional theories on gender stereotypes, the results suggest that female leaders are not subjected to harsher punishment when revealed as overconfident compared to male leaders. These findings offer valuable insight into the interpersonal consequences of overconfidence among high-level leaders and shed light on the influence of gender.

Table of Contents

1 Introduction

In 2001, when the once highly regarded Wall Street Giant, Enron, crumbled almost overnight, it sent shockwaves through the world. The sudden and dramatic downfall left countless individuals who had lost billions of dollars angry and disillusioned. It is a fair bet that many would place a lot of blame for such corporate disasters on leader *overconfidence* (Cheng et al., 2021). The Enron scandal in 2001 serves as just one among numerous examples of overconfidence. There is a prevailing consensus among the public that overconfidence is harmful (Moore & Bazerman, 2022). Accordingly, overconfident individuals might be penalized for their perceived hubris, as people accord lower social status to individuals who hinder the group's success (Ridgeway & Diekema, 1989).

This taken into consideration, the high prevalence of overconfidence among corporate executives is rather surprising (Heaton, 2002; Lee et al., 2017; Reyes et al., 2022). Take the example of Elon Musk, the eccentric CEO of Tesla. He is widely recognized for his unwavering self-assurance. While his arrogance has occasionally caused anxiety among shareholders, such as when a single tweet led to a $14 billion drop in Tesla's value in 2020, it appears to have paid off as he is currently the second wealthiest man on earth, with an estimated net worth of almost $180 billion (Forbes, 2023). The widespread occurrence of overconfidence among corporate executives suggests that it may indeed confer some social advantages (Kennedy et al., 2013). Previous research on the interpersonal effects of overconfidence has yielded contradictory results. Two opposing lines of research exist, one showing that overconfidence yields social advantages (Anderson & Kilduff, 2009; Anderson et al., 2012; Kennedy et al., 2013; Price & Stone, 2004; Ronay et al., 2019) and one showing that it does not (Paulhus, 1998; Sah et al., 2013; Tenney et al. 2007, 2008; Tenney & Spellman, 2011; Vullioud et al., 2017). However, a great deal of research has been conducted without an accurate benchmark (Moore & Schatz, 2017). The present study follows the experimental paradigm of Tenney et al. (2007, 2008), Tenney and Spellman (2011), Kennedy et al. (2013), and Sah et al. (2013), wherein overconfidence is manipulated, as opposed to naturally occurring overconfidence, and respondents have the opportunity to easily assess accuracy. In the present study, we seek to answer the following question:

(1) *Is it more beneficial for leaders to be calibrated than overconfident when actual abilities are revealed to others?*

A solid body of research indicates that men tend to exhibit higher levels of confidence compared to females (e.g., Huang & Krisgen, 2013). Conventional theories about gender and leadership posit that females that behave counterstereotypically are likely to encounter negative repercussions (e.g., Eagly & Karau, 2002; Rudman & Phelan, 2008). When revealed as *over*confident, this backlash is likely to be amplified as research indicates that mistakes are particularly damaging for females in gender-incongruent occupations, such as in senior leadership positions (e.g., Brescoll et al., 2010). Hence, we also seek to answer the following question:

(2) *Are female leaders penalized harder when revealed as overconfident than equally overconfident male leaders?*

The present study seeks to contribute to the debate about the interpersonal consequences of overconfidence. By utilizing simulation experiments, we manipulated both confidence levels and actual abilities. This approach was employed with the objective of establishing causal priority. A substantial body of research has examined the interpersonal consequences of overconfidence in various contexts, such as hiring simulations and witness depositions. However, we add to the literature by assessing overconfidence specifically within the realm of high-level leadership. To our knowledge, no studies have yet found a gender effect on the interpersonal consequences of overconfidence. Thus, we hope to bridge this gap in the literature.

The present paper is structured as follows: Firstly, we review existing literature on overconfidence and gender stereotypes. Next, we outline our research methodology, explaining the data collection and analysis procedures. We then present the results of our data analysis. Following this, we engage in a discussion to interpret the results within the context of our research question. Subsequently, we critically address the study's limitations, identifying potential constraints and areas for further investigation. Lastly, we briefly explore the practical implications of our research and conclude by summarizing the main findings.

2 Literature Review

2.1   Overconfidence

2.1.1   Defining Overconfidence

Overconfidence is a classic topic in the research field of judgment and decision making and is considered one of the most common and impactful biases in human thinking (Plous, 1993). In short, overconfidence can be defined as an inaccurate and exaggerated perception of one's abilities or knowledge (Ronay et al., 2019). The literature has defined overconfidence in three distinct ways. The first form of overconfidence is *overestimation,* thinking that you are better than you are (e.g., thinking that you answered seven of ten questions correctly when you only got two). The second form is *overplacement,* the exaggerated belief that you are better than others (e.g., thinking your score on the test is ranked top in the class when you scored second to last). The third form of overconfidence is *overprecision*, the excessive faith that you know the truth (e.g., being 100 % convinced that you got seven questions right when you did not) (Moore & Schatz, 2017). The present study captures overconfidence in terms of overestimation – the tendency to believe that you are better than you actually are (Moore & Schatz, 2017). We opted for this particular definition as it has been extensively studied in the context of CEO overconfidence and impact expectations regarding financial returns (e.g., Engelen et al., 2015; Ho et al., 2016; Malmendier & Tate, 2008). Thus, overestimation is arguably most applicable to the present study design. It is worth noting that scholars sometimes use constructs such as confidence, self-assurance, arrogance, self-promotion, or managerial hubris interchangeably with overconfidence (Hayward & Hambrick, 1997; Hayward et al., 2006; Li & Tang, 2010). Although there is indeed an overlap, they differ as overconfidence, by default, entails an exaggerated perception of knowledge/abilities, whereas these constructs do not necessarily involve a misjudgment of one's abilities, though they often do (Reyes et al., 2022).

Initially, research on overconfidence focused on either documenting its occurrence (e.g., Fischhoff et al., 1977), harmful effects (e.g., Camerer & Lovallo, 1999; Johnson, 2004; Malmendier & Tate, 2005), or finding ways to reduce it (e.g., Arkes et al., 1987). For instance, a vast majority of people say they are above average when assessing various skills and abilities, such as driving abilities (Svenson, 1981), performance on exams (Shepperd et al., 1996), and income

prospects (Weinstein, 1980), although, of course, only half can be. Even travelers overestimate how much fun they will have on vacation (Mitchell et al., 1997). Indeed, psychology scholars provide strong evidence that supports the existence of this cognitive bias in various domains, such as politics, financial markets, and leadership (Moore & Healy, 2008). Yet, overconfidence can lead to poor decision making that can negatively affect individuals, organizations, and even nations. For instance, overconfidence can cause entrepreneurs to take excessive risks when launching new ventures (Camerer & Lovallo, 1999), CEOs to engage in too many acquisitions (Malmendier & Tate, 2005), and nations to initiate military confrontations (Johnson, 2004). Consequently, overconfidence has been offered as an explanation for phenomena such as bankruptcy, stock market bubbles, and even wars (Moore & Healy, 2008). According to Plous (1993), *"No problem in judgment and decision making is more prevalent and more potentially catastrophic than overconfidence"* (p. 217). Previous research offers a pessimistic assessment of the effectiveness of debiasing tools, such as collaboration, warnings, feedback, and expertise. Research indicates that these tools do not consistently reduce overconfidence and may even exacerbate it (Meikle et al., 2016).

2.1.2 Interpersonal Consequences of (Over)Confidence

More recent studies have examined the interpersonal consequences of overconfidence. That is, how people perceive overconfident individuals. A large body of research suggests that the display of confidence, a high degree of certainty in one's judgment/ability, increases interpersonal influence in beneficial ways (e.g., Cramer et al., 2011; Semmler et al., 2012; Tenney & Spellman, 2011). This is illustrated in an experiment by Zarnoth and Sniezek (1997), where participants were asked to respond to various kinds of problems, first individually and then in groups. They found that group responses tended to match the individual response of the most confident group members, regardless of whether the response was correct. This is consistent with findings by Sniezek and Van Swol (2001), who discovered that mock judges were more likely to take advice that was expressed with high confidence, and Whitley and Greenberg (1986), who conducted a series of experiments using mock jurors and found that confidence of eyewitness testimony is the single best predictor of perceived credibility.

Price and Stone (2004) argue that such findings can be attributed to a *confidence heuristic.* That is, people use confidence as a cue for competence and assume that confident individuals are most likely to be correct. Importantly, perceived competence is a primary and robust predictor of social status in groups and generates respect, prominence, influence, and credibility (Oh et al., 2020). This is consistent with the *presumption of calibration hypothesis*, which states that confidence is compelling because, in the absence of evidence to the contrary, people assume that others have good self-insight. That is, the default is to assume that a person's confidence is a good indicator of accuracy (Tenney & Spellman, 2011). Given that we presume calibration, then confidence does imply accuracy (Sah et al., 2013). In the absence of information, using a confidence heuristic is not unreasonable (Price & Stone, 2004). For instance, a meta-analysis by Miller et al. (2015) revealed an improvement in accuracy associated with greater confidence in clinical decision making. This suggests that confidence is, to *some* extent, calibrated with accuracy. However, the effect size was rather small ($r = .15$) (Miller et al., 2015). That is, confidence accounts for only about 2 % of the variance in judgment accuracy ($r^2 = 0.0225$). If people were able to assess the accuracy of their own judgments appropriately, the aggregated effect size would be significantly larger.

As it turns out, people appear to be poorly calibrated, and confidence is often inflated (Hardman, 2009, p. 94). As overconfidence is a genuinely flawed perception, it gives rise to the very same behavioral signals as confidence, such as strong opinions, self-assurance, and defined ideas (Ronay et al., 2019). That is, people cannot easily distinguish between justifiable confidence and overconfidence in the absence of objective data as actual task ability is hidden within them, yet they exhibit the same behaviors (e.g., Anderson & Kilduff, 2009; Vrij et al., 2000). Hence, overconfidence can be expected to yield the same social benefits as confidence. Ronay et al. (2019) provide evidence that overconfidence gives an appearance of knowledge or skill that provide benefits in the context of selecting leaders. Using a multi-method approach, including a field study, that assessed the relationship between overconfidence and perceived leadership suitability, they found overconfidence to predict hiring recommendations for advertised leadership positions and increase perceived leadership potential. This is consistent with findings by Anderson and Kilduff (2009), who conducted two laboratory experiments wherein groups worked together on various tasks. They

found that overconfident group members were rated as more competent, even after controlling for actual abilities. Similarly, a series of studies by Anderson et al. (2012), using naturalistic and experimental designs, found that overconfident individuals are perceived as more competent by group members. Overconfident individuals attained higher status in the group in terms of respect, prominence, and influence. This indicates that it is sufficient to be perceived as possessing the competence and/or skill to attain social status, regardless of whether you actually possess these characteristics. This is a core idea of the *status-enhancement theory* of overconfidence, which suggests that the prospect of obtaining higher social status encourages displays of high confidence (Kennedy et al., 2013). Overall, these studies suggest that people are swayed by confidence, even when unjustified.

Overconfidence is Beneficial

The studies by Ronay et al. (2019), Anderson and Kilduff (2009), and Anderson et al. (2012) all have in common that they used naturally occurring overconfidence, thus precluding strong inferences of causality. Additionally, if targets have a limited chance of being recognized as overconfident, overconfidence is unlikely to elicit any negative perceptions. In real-life settings, group members may be able to more accurately discern each other's competence when working together over time (e.g., Vullioud et al., 2017). This is particularly the case for high-profile leaders whose performance will be subject to close scrutiny (Rosette & Tost, 2010). If so, penalties may eventually outweigh the initial social benefits overconfidence yield. Then, what happens when overconfidence is revealed to others?

Price and Stone (2004) conducted experiments where college students evaluated two fictional financial advisors, Advisor Green and Advisor Brown, who judged the probability that stocks would increase in value. One of the advisors was well-calibrated (the moderate advisor), and one was overconfident (the extreme advisor). Participants received objective performance feedback in terms of a statement of whether or not the stock actually increased or decreased in price. Price and Stone (2004) found that in all experiments, participants displayed a tendency to prefer the overconfident advisor to the better-calibrated one. This is surprising as the moderate advisor was, in fact, more accurate and could be expected to be preferred by participants. Laboratory studies conducted by Kennedy et al. (2013) yielded similar findings. Across three experiments, using

both naturally occurring overconfidence and trained actors, they found that group members did not react negatively to individuals revealed as overconfident. On the contrary, they still viewed them positively. Overconfident participants were accorded higher status when others were unaware of actual task performance. Yet, they were not penalized with lower status in the group when others gained clear, objective information about individual task performance. This taken into consideration, there is indeed evidence that suggests that, on balance, the social benefits of overconfidence outweigh its social costs even after overconfidence is revealed to others.

Overconfidence is Not Beneficial

However, the findings are inconclusive. Common intuition suggests that individuals are likely to respond unfavorably upon realizing someone's overconfidence (Kennedy et al., 2013). A number of opposing studies suggest that overconfidence is penalized by others upon discovery. For instance, Paulhus (1998) conducted two longitudinal studies where groups of 4-7 participants met seven times, rating their perception of one another after the first and final meeting. Group members with unrealistically positive self-evaluations made positive impressions at the first meeting, being perceived as agreeable, competent, and well-adjusted. However, they were rated negatively after the final meeting. This suggests that group members are indeed better at detecting overconfidence after several encounters.

Tenney et al. (2008) argue that people are initially positive to individuals who appear highly confident because they believe the confidence is grounded in reality, consistent with the presumption of calibration hypothesis. Across two experiments where participants acted as jurors and had to choose between opposing witnesses to a car accident, they found perceived credibility to ultimately depend on good calibration. Similarly, experiments by Tenney and Spellman (2011) revealed that fictitious applicants for a spot on a swim team or daycare who were confident about their qualities were initially evaluated by participants as stronger candidates than more cautious applicants. However, when participants learned that confidence was unwarranted, high confidence backfired. In fact, the more cautious candidate was preferred over the confident candidate, even though they were equally good. Furthermore, Tenney et al. (2007) conducted experiments where participants read written trial summaries of crimes that allegedly occurred on a college campus. They found that the erroneous witness

testimony damaged perceived credibility more when the witness was confident about the testimony. This is consistent with laboratory studies conducted by Sah et al. (2013) using fictional advisors, who found that while confidence was beneficial for accurate advisors, inaccurate yet confident advisors (i.e., overconfident advisors) received low credibility ratings. More recently, Vullioud et al. (2017) conducted a series of experiments in classrooms and online, in which participants received identical advice from fictional senders who differed in terms of confidence and competence. When it became evident that the advice was misguided, the sender displaying exaggerated confidence experienced a greater decline in reputation compared to the competent sender. Consequently, Vullioud et al. (2017) concluded that having an inflated sense of confidence can lead to adverse consequences and damage an individual's reputation. This taken into account, a substantial body of research suggests that people tend to reassess their initial positive opinion when they suspect a lack of self-knowledge in others. Overconfident individuals may lose face and appear less competent and likable than well-calibrated individuals (Tenney & Spellman, 2011). Based on these studies, it seems that although confidence may have social advantages, it is more advantageous to demonstrate accurate self-assessment of one's knowledge in the long run, as overconfidence is likely to backfire.

Overall, there is a high degree of consensus in the academic literature regarding the social benefits associated with confidence. Yet, the literature is highly divided regarding the interpersonal consequences when confidence is revealed as unwarranted. Drawing upon the theory presented above, we have developed the following hypotheses:

H1: We expect a CEO who displays high confidence to be evaluated as more competent than a modest CEO

H2: After receiving objective performance data, we expect a well-calibrated CEO to be rewarded with higher competence ratings

H3: After receiving objective performance data, we expect an overconfident CEO to be penalized with lower competence ratings

H4: All else equal, we expect that calibration trumps overconfidence when objective performance data is revealed to participants

Note that hypotheses H3-H4 are more tentative than H1-H2, as previous research has yielded conflicting findings.

2.2   Competence and Warmth

The dependent variables in the present study are perceptions of *competence* and *warmth*. Research has firmly established that perceived competence and warmth, as proposed by the *Stereotype Content Model*, are two universal dimensions of human social cognition (Fiske et al., 2007). Social perception, like all types of perception, is a reflection of evolutionary pressures. In interactions with members of their own species, social animals are required to quickly assess whether the other individual intends good or ill (friend or foe) (i.e., the warmth dimension) and whether they possess the capability to carry out those intentions (i.e., the competence dimension) (Fiske et al., 2007). Although the literature has employed different labels, such as morality and competence (Phalet & Poppe, 1997), a high degree of consensus prevails that competence and warmth can universally capture social perception (Fiske, 2018). In fact, the basic dimensions of warmth and competence have been found to explain 82 % of the variation in perceptions of everyday social behaviors (Wojciszke et al., 1997).

Interestingly, research indicates a negative dynamic relationship between perceptions of competence and warmth – meaning that an apparent surplus of one trait implies a deficiency of the other (Judd et al., 2005). In fact, experiments show that the more competent the target is, the less warm the target is believed to be, and vice versa (Kevryn et al., 2009). Importantly, in organizational contexts, competence judgments have been found to have a greater impact than warmth judgments (Cuddy et al., 2011). This has been particularly well-documented in the leadership field (e.g., Cejka & Eagly, 1999; Eagly, 1987; Eagly & Karau, 2002; Glick et al., 1995; Heilman, 1983; Prentice & Carranza, 2002). Corporate executives are generally depicted as high in competence but low in warmth (e.g., "business leaders are rich and smart, but they are arrogant and calculating"). This can partly be attributed to the fact that most leaders are more concerned about creating and maintaining an aura of competence than an aura of warmth (Cuddy et al., 2011). Another explanation is that individuals characterized as warm but incompetent are underrepresented in high-level leadership positions (Fiske et al.,

2007). Consequently, our primary focus in the present study is on perceptions of competence. As our experimental design deliberately avoids manipulating perceptions of warmth, we anticipate a minimal impact on this variable. Thus, we consider perceptions of warmth as a secondary outcome variable. Based on the literature presented, we developed the following hypothesis:

H5: We expect evaluations of warmth to remain largely unaffected when objective performance data is revealed to participants

2.3   Gender Differences

Gender stereotype research indicates that unwarranted confidence may affect perceptions of female and male leaders somewhat differently. *Social role theory* suggests that stereotypes of males and females are based on characteristics related to communal and agentic attributes (Eagly, 1987). *Communal* attributes, often associated with females, include being caring, supportive, empathetic, and gentle. On the other hand, *agentic* attributes, often associated more with males, include being ambitious, competent, dominant, independent, and self-confident (e.g., Eagly, 1987; Rudman & Phelan, 2008; Williams & Best, 1990). To draw parallels to the competence and warmth dimensions, stereotypes depict females as generally warmer than males and males as generally more competent than females (Rudman & Glick, 2001). Consequently, early research demonstrated a *think manager-think male* effect (Schein, 1973; 1975), as there is a perceived lack of fit between attributes associated with females and attributes associated with leaders (Heilman, 1983, 1995, 2001).

Gender stereotypes are not only descriptive; they are also prescriptive. That is, they signify not only differences in how females and males actually are but also norms regarding behaviors that are suitable for each—about how females and males should be (Burgess & Borgida, 1999; Eagly, 1987; Terborg, 1977), and importantly how they should not be (e.g., Fiske & Stevens, 1993; Prentice & Carranza, 2002). To incorporate the prescriptive aspect of gender stereotypes, the *role congruity theory* was later proposed as an extension of the social role theory (Eagly & Karau, 2002). According to role congruity theory, two forms of prejudice against female leaders exist. The first is the perception that females are less suitable for leadership roles than males. The second is that when a female behaves in ways that are expected of a leader, she is often evaluated less favorably

than a male would be in the same situation. According to this theory, females who behave in ways that are traditionally associated with femininity are seen as incompatible with leadership roles, while those who exhibit behaviors that are expected of leaders are viewed as not congruent with being a female. This ultimately leads to a "damned if they do and doomed if they don't"- kind of situation (Catalyst, 2007, p. 1).

A prevalent finding in psychological research is that males are more overconfident than females (e.g., Barber & Odeon, 2016; Bhandari & Deaves, 2006; Huang & Kisgen, 2013; Lundeberg et al., 1994). Compared to males, females are expected to downplay their own achievements and emphasize the contributions of others (Budworth & Mann, 2010). Consistent with role congruity theory, females that express confidence face the risk of being perceived as boastful, arrogant, and braggy, which ultimately may decrease likeability (Daubman & Sigall, 1997). This is unfortunate as female leaders are prone to encountering *the double bind*, wherein female leaders, more than their male counterparts, face the need to be warm (i.e., gender societal norms) as well as competent (i.e., societal leadership norms) to avoid backlash effects (Trzebiatowski et al., 2023). Backlash effects refer to social and economic reprisals for behaving counterstereotypically (Rudman et al., 1998). In fact, there has been found evidence of a so-called *glass cliff,* implying that females in top-level leadership positions are more susceptible than males to fall from their position (Ryan & Haslam, 2005). The introduction of the metaphor was prompted by the observation that females are more prone to receiving promotions to high-level leadership positions during periods of crisis. However, there is evidence that failure is more likely to have a detrimental impact if the leader is a female than a male (Ryan & Haslam, 2005). For instance, an experiment conducted by Brescoll et al. (2010) revealed that female leaders who make one single mistake are accorded less status and perceived as less competent than their male counterparts. This is consistent with observations by Williams and Dempsey (2014), who reported that females' reputations are far more tenuous and fragile than males' reputations. Thus, if a failure occurs, females are more likely to be singled out for criticism and blame (Trzebiatowski et al., 2023).

A plethora of empirical investigations in real-world settings demonstrate criterion-related validity of the aforementioned cited studies. For instance, Albanesi et al. (2015) found that female CEOs suffer more severe consequences

when firms lose money in terms of stock-based pay. Kennedy et al. (2016) reported that female managers receive harder penalties for committing ethical violations. Additionally, research by Gupta et al. (2018) revealed that female CEOs are more likely than male CEOs to face shareholder activism. An older meta-analysis by Swim and Sanna (1996) can shed light on these findings. They reported that a female's failure is more likely to be attributed to low ability (i.e., internal factors), whereas a male's failure is more likely to be attributed to bad luck (i.e., external factors). Although these findings probably reflect implicit biases, Fisk and Overton (2019) discovered that laypeople also expect female leaders to be punished more harshly for failure than otherwise similar males, suggesting that people are aware of this gendered discrimination (Fisk & Overton, 2019). Overall, previous research clearly indicates gendered sanctions for leadership failure, suggesting that female leaders stand more to lose relative to their male counterparts.

Based on the theory presented, we expect CEO gender to moderate the strength and/or direction between confidence level and interpersonal evaluations. We propose the following hypothesis:

H6: We expect a female CEO to be penalized harder than a male CEO when revealed as overconfident

2.4   The Present Research

The present study seeks to contribute to the debate about the interpersonal effects of overconfidence by examining if it is most beneficial for leaders to be overconfident or well-calibrated when actual abilities are disclosed. By providing objective and unambiguous performance feedback, we are able to rule out the possibility that participants do not register the targets' overconfidence. A considerable body of research on interpersonal consequences of overconfidence has used hiring simulations (e.g., Tenney & Spellman, 2011; Ronay et al., 2019), fictitious advisors (e.g., Sah et al., 2013; Price & Stone, 2004; Vullioud et al., 2017), witness depositions (e.g., Tenney et al., 2007; Tenney et al., 2008) and task performance in groups (e.g., Paulhus, 1998; Kennedy et al., 2013; Anderson & Kilduff, 2009; Anderson et al., 2012). The present study differs from the studies cited above, as overconfidence is assessed in the context of high-level leadership. Although the association between leader overconfidence and firm performance

has been extensively studied (e.g., Chen et al., 2014; Kim & Jang, 2021; Malmendier & Tate, 2008; Weng & Yamakawa, 2022; Lee et al., 2023), no studies to our knowledge have examined interpersonal effects of overconfidence among high-level leaders. Additionally, we examine whether gender impacts the strength and/or direction of the relationship between leader overconfidence and perceived competence and warmth. Although perceptions of females in high-level leadership positions have received substantial attention in scientific realms, no studies to our knowledge have examined how unwarranted confidence is perceived depending on the target's gender. Hence, little empirical and theoretical knowledge exists in this domain.

Table 1

Conceptual Framework
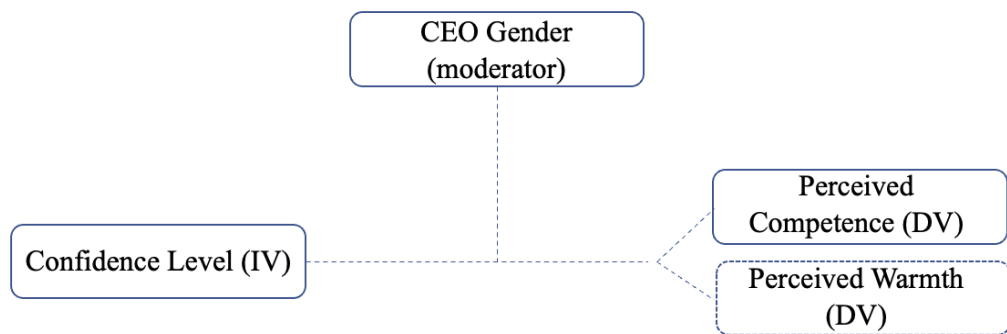
3 Method

3.1 Participants

The survey was distributed to employees of two prominent Norwegian corporations, Gjensidige and Veidekke. Additionally, we distributed the survey link through social media channels (i.e., Facebook, LinkedIn) and encouraged our extended network to recruit participants among their friends and acquaintances as we wanted to access the opinion of a broad range of Norwegians. In total, 561 people responded to the survey. After excluding participants who had substantial missing data or failed to respond correctly to one or more control questions, the final dataset consisted of 454 participants (203 females, 251 males). An *a priori*

sample size of 454 was computed for a mixed ANOVA with a power of 0.8, indicating a sufficient likelihood of detecting meaningful effects (Anderson et al., 2017). Most participants were employees in Gjensidige (35 %) and Veidekke (28,4 %), constituting about 63 % of the participants. The participant's ages ranged from 19 to 91 years (M = 42.32 SD = 12.99). Most participants reported having at least some higher education (75.3 %). Furthermore, 40.7 % reported being in a leadership position. It is worth mentioning that we used a non-random sampling method and recruited participants that were easily accessible to us (i.e., convenience sample), which may potentially hamper the generalizability of our findings (Bordens & Abbott, 2014, p. 161). Caution should be exercised when extrapolating the results to the wider population. This will be further addressed in limitations.

3.2 Materials

3.2.1 Measurement of Competence and Warmth

Data were collected using a questionnaire on the web-based Qualtrics XM platform (https://www.qualtrics.com/). The scales used in the study to assess the primary outcome measure, competence and warmth, were derived from Halkias and Diamantopoulos (2020). They identified a set of items that capture perceptions of competence and warmth with strong psychometric properties. They recognized six items that measure competence consistently: capable, competent, efficient, industrious, skillful, and intelligent, and five items that measure warmth consistently: friendly, kind, likable, nice, and warm. Cronbach's alpha for the measures derived from Halkias and Diamantopoulos (2020) was high for each subscale, α_{warmth} = .93 and $\alpha_{competence}$= = .92, indicating sufficient internal consistency (Bordens & Abbott, 2014, p. 126). In the present study, we obtained similarly high levels of reliability for these measures, with Cronbach's alpha values of $\alpha_{warmth(1)}$ = .92 and $\alpha_{warmth(2)}$ = .95 and $\alpha_{competence(1)}$ = .89 and $\alpha_{competence(2)}$ =.92. An example item is "Based on first impression, to what degree do you perceive [name CEO] as intelligent?". A 5-point Likert scale was used to evaluate the CEOs, ranging from 1 (very small extent) to 5 (very large extent). As the survey was administered to two Norwegian companies, we translated the measurement items from the original language (English) to Norwegian (see Appendix A). This led us to exclude three items from the scale (i.e., likable, capable, industrious) as the adjectives for each construct appeared too similar in Norwegian. Although

translating the measurement items may pose a threat to construct validity (Bordens & Abbott, 2014, p. 129), enabling participants to respond to questionnaires in their first language is important to minimize misinterpretation and mitigate potential threats to the reliability of the results (Harkness & Schoua-Glusberg, 1998).

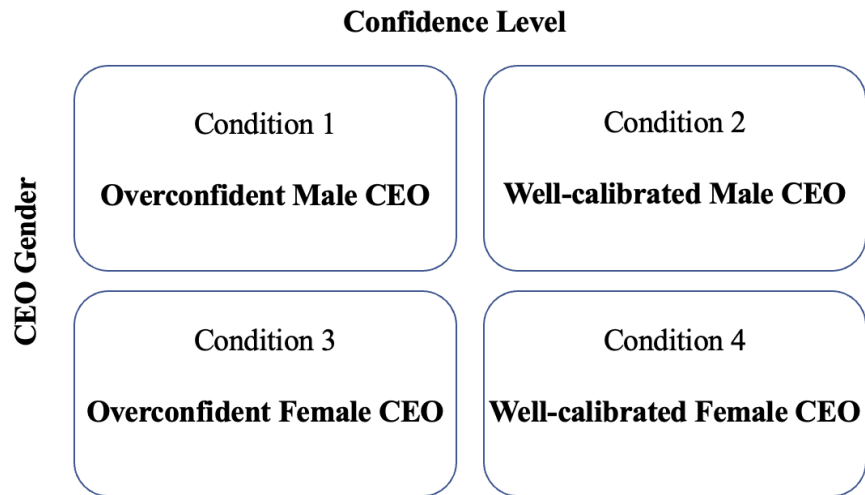### 3.2.2 Attitude Scale for Leader Optimism and Certainty

An additional measure was included in the study to control for individual differences, as some may have a preference or affinity for overconfidence compared to others. In order to measure opinions regarding the appropriate level of leader optimism and certainty, we derived an attitude scale from Armor et al. (2008). Two items were utilized to measure attitudes to leader optimism/certainty on a scale from -3 (very pessimistic/uncertain) through 0 (accurate) to +3 (very optimistic/ certain) (see Appendix B).

3.3 Research Design

The experiment used a 2 (between-subjects: overconfident CEO vs. well-calibrated CEO) x 2 (between-subjects: female CEO vs. male CEO) x 2 (within-subjects: Phase 1 [before objective performance feedback] vs. Phase 2 [after objective performance feedback]) mixed factorial design. That is, the independent variables in the present study are confidence level, CEO gender, and time of measurement [before vs. after performance feedback]. Our dependent variables are perceived competence and warmth. We used an experimental design as we aimed to draw causal inferences about the impact of confidence level on perceptions of competence and warmth and whether gender moderates this relationship (Bell et al., 2015, p. 51).

Table 2

The Experimental Conditions

Confidence Level

	Condition 1 **Overconfident Male CEO**	Condition 2 **Well-calibrated Male CEO**
CEO Gender	Condition 3 **Overconfident Female CEO**	Condition 4 **Well-calibrated Female CEO**

3.4   Procedure

3.4.1   Pilot Study

In order to mitigate potential threats to validity, we conducted a pilot study (Van Teijlingen & Hundley, 2002). A sample of 20 participants was recruited to help identify any potential issues that could arise. Upon completing the survey, participants were asked to report any instances of confusion or misunderstanding. Some participants expressed concern about the similarity and vagueness of a few adjectives used to measure competence and warmth. Thus, we eliminated three items to improve the clarity and specificity of the measurement. Ensuring that the measurement items are clear, specific, and free from ambiguity can enhance the internal consistency and reliability of the scale (Bordens & Abbott, 2014, p. 126) Additionally, the pilot test revealed that the objective performance feedback contained excessive information and was difficult to read on mobile devices. In order to make the objective performance feedback more comprehensible and mobile-friendly, we divided the information into several pieces and removed all unnecessary words and numbers. The pilot study allowed us to identify issues and ensure that the experimental conditions seemed realistic and that the manipulation worked as intended. To avoid data contamination, the data collected from the pilot study (N=20) was excluded from the main results, and new data was not collected from these participants (Van Teijlingen & Hundley, 2002).

3.4.2   Study Procedure

After giving their informed consent and responding to some demographic questions (see Appendix C), participants were randomly allocated to one of four different experimental conditions; (1) overconfident male CEO, (2) well-calibrated male CEO, (3) overconfident female CEO, or (4) well-calibrated female CEO. We attained a nearly equivalent number of participants for each of the four experimental conditions (see Appendix D). The participants were not informed which condition they were assigned to nor that other experimental conditions existed. Participants were merely told that the purpose of the study was to investigate how leaders are perceived. To operationalize overconfidence, we created a podcast (duration of approximately one minute) depicting a new CEO of a fictitious Norwegian bank talking about his/her expectations of how the company would perform following an acquisition. In the audio clips, the overconfident CEO expressed high confidence regarding the financial expectation of the acquisition and used statements like, "I *expect* to reach a market share in the private market of 30 %.". The well-calibrated CEO expressed more modest expectations regarding the performance of the acquisition and used statements like "I *hope* to reach a market share in the private market of 25 %.*". Beyond this, experimental conditions were kept as identical as possible to reduce noise and enhance internal validity. Except for the estimates and subtle differences in confidence expressions (e.g., expect versus hope), the manuscripts were identical (see Appendix E). We used audio clips rather than video clips to limit the impact of biases that could pose a threat to internal validity (e.g., liking effect, body language).

To manipulate gender, the voices of the CEOs were recorded by two actors, a middle-aged female and male with the same dialects and the same amount of leadership experience. The actors were instructed to use the same tone of voice and emphatic expressions. In addition, images of either a male or female in business attire were attached, featuring the CEOs with the same posture and roughly the same level of attractiveness (see Appendix F). The same image was used in both the well-calibrated and overconfident conditions. After listening to the brief audio clip, participants were asked to rate how they perceived the CEO in terms of competence and warmth. The items were presented in a randomized order to minimize potential order effects.

At the next stage of the experiment, participants received excerpts from a fictitious annual report two years following the acquisition (see Appendix G). The

manuscript and fictitious annual reports were reviewed and validated by an experienced professional in mergers and acquisitions from a prominent securities firm in Norway. This measure was taken to ensure that the information provided was realistic and credible. Importantly, the actual performance was identical across conditions, and the acquisition had been fairly successful in terms of revenue and market share. Whereas the well-calibrated CEOs gave estimations fairly equivalent to the annual report, the overconfident CEOs had substantially overestimated the impact of the acquisition (see Table 3). After participants were provided with objective performance feedback in terms of achieved revenue and market share, participants were asked three easy control questions to verify that they had read the brief excerpts, for example: "Is the market share in 2022 higher or lower than expected?". With the newly acquired information, participants revised their initial perception of the CEO in terms of competence and warmth using the same scale. The participants were not able to modify their initial evaluations. After the second evaluation, participants responded to two general questions regarding their opinions on the appropriate level of optimism and certainty leaders should display when engaging with stakeholders.

Table 3

CEO Projections versus Actual Performance

	Overconfident CEO			Well-calibrated CEO		
	Estimated	Achieved	Difference	Estimated	Achieved	Difference
Revenue	50 %	21 %	- 29 %	20 %	21 %	+ 1 %
Market Share	10 %	5 %	- 5 %	5 %	5 %	0 %

Note. The overconfident and well-calibrated CEO had the same starting point prior to the acquisition

3.5   Data Diagnostics

The data were exported from Qualtrics directly into IBM SPSS v. 29. Parametric inferential tests are based on several assumptions, which should be considered prior to running ANOVAs (Bordens & Abbott, 2014, p. 433). Firstly,

the observations have to be independent. This assumption is justified as we used proper randomization to experimental treatment (e.g., Casler, 2015; Piepho et al., 2013). Second, the dependent variable has to be continuous. We measured dependent variables using 5-point Likert scales. Although such scales could be seen as ordinal, we follow the common practice of treating them as continuous. Third, the sampling distribution of the mean should be normal. As the histograms approximate a bell curve (see Appendix H), and Q-Q Plots show that the data points are close to the diagonal line (see Appendix I), this assumption is arguably met (Emerson, 2020). Fourth, the error variances of all data points of the dependent variable should be equal or homogenous throughout the sample (i.e., homoscedasticity). Levene's Test show that for each of the dependent variables (except Competence Phase 2), the null hypothesis cannot be rejected, with a non-significant result ($p > .05$), indicating that equal group variance exists (i.e., meeting the assumption of homogeneity) (see Appendix J). However, the null hypothesis for Competence Phase 2 ($p = 0.02$) is rejected, indicating unequal variance. Levene's test relies heavily on p- values and, consequently, the sizes of the samples (Kozak & Piepho, 2018) and will yield a lower p-value for larger samples than smaller ones. Moderate deviations in variance are unlikely to have a substantial impact on our dataset as the ANOVA is widely recognized for its robustness in handling violations of this assumption when the sample size is equivalent across groups, as is the case in the present study (Kozak & Piepho, 2018).

4 Results

4.1 Competence

Phase 1 Ratings. In Phase 1, when participants did not have information to verify predictions, a 2 x 2 ANOVA with confidence level and gender as between-subjects factors demonstrated no main effect of confidence level, $F(1, 450) = 1.270$, $p = .260$, $\eta_p^2 = .003$, failing to support H1. The confident CEO was not evaluated as more competent than the modest CEO prior to performance feedback. However, a significant main effect of CEO gender was found, $F(1, 450) = 54.301$, $p < .001$, $\eta_p^2 = .108$. Participants rated the female CEO ($M = 3.93$, $SD = .70$) as more competent than the male CEO ($M = 3.42$, $SD = .76$). No interaction effects between CEO gender and confidence level were found $F(1, 450) = .401$, $p = .527$, $\eta_p^2 = .001$. That is, findings indicate that CEO gender did not moderate the

relationship between confidence levels and perceived competence prior to performance feedback.

Table 3

Mean Ratings (and Standard Deviations) of Competence and Warmth in Phase 1 and Phase 2.

	Perception of competence		Perception of warmth		N	
	Female CEO *M (SD)*	Male CEO *M (SD)*	Female CEO *M (SD)*	Male CEO *M (SD)*	Female CEO	Male CEO
Modest (Phase 1)	3.95 (.66)	3.48 (.77)	3.11 (.81)	2.75 (.85)	115	110
Confident (Phase 1)	3.91 (.74)	3.36 (.76)	2.94 (.86)	2.44 (.79)	116	113
Total	3.93 (.70)	3.42 (.76)	3.02 (.84)	2.60 (.83)	231	223
Well-calibrated (Phase 2)	4.17 (.66)	3.85 (.70)	3.18 (.82)	2.78 (.90)	115	110
Overconfident (Phase 2)	3.45 (.81)	2.84 (.70)	2.89 (.87)	2.44 (.82)	116	113
Total	3.81 (.82)	3.34 (.87)	3.03 (.86)	2.61 (.87)	231	223

Note. Means (and standard deviations) of competence and warmth ratings as a function of CEO gender, confidence level, and time of measurement (Phase 1 and Phase 2). *N*=454

Table 4
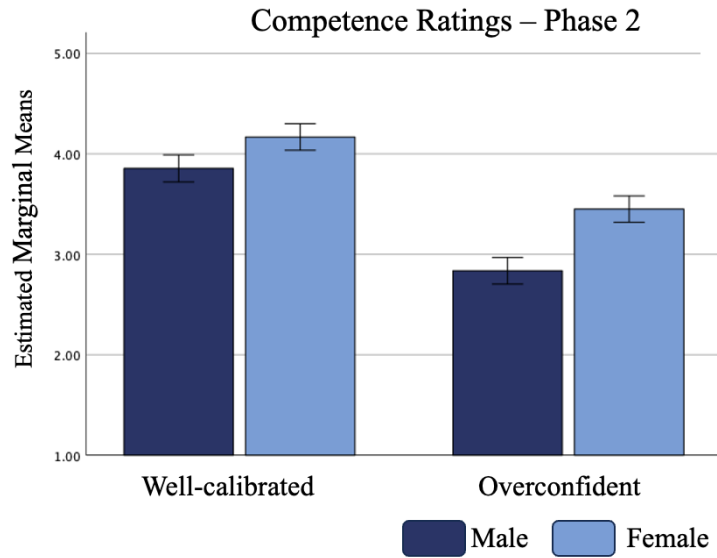*Competence Ratings – Phase 1*

Competence Ratings – Phase 1
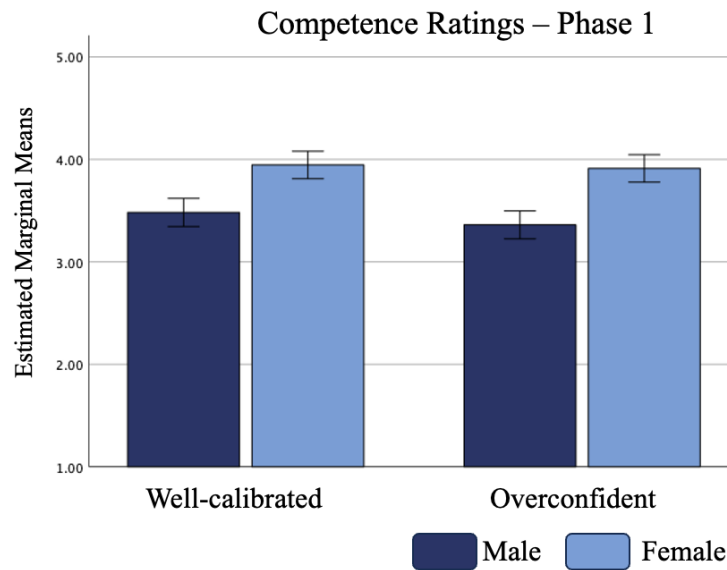
Note. Competence ratings in Phase 1, depending on the confidence level and CEO gender. Error bars: ± 1 SEM.

Phase 2 Ratings. After objective performance was revealed, a 2 x 2 ANOVA showed a main effect of confidence level, $F(1, 450) = 165.108$, $p < .001$, $\eta_p^2 = .268$. Summarizing across conditions, participants rated the well-calibrated CEO ($M = 4.01$, $SD = .70$) as considerably more competent than the overconfident CEO ($M = 3.15$, $SD = .81$) in Phase 2. This finding is consistent with H4, wherein we expected the well-calibrated CEO to be evaluated more favorably than the overconfident CEO after receiving performance feedback. As with Phase 1, the female CEO was evaluated as significantly more competent than the male CEO in Phase 2, $F(1, 450) = 47.121$, $p < .001$, $\eta_p^2 = .095$. Furthermore, we found an interaction effect between CEO gender and confidence level, $F(1, 450) = 4.978$, $p = .026$, $\eta_p^2 = .011$. Pairwise comparisons (Bonferroni- corrected) showed that the discrepancy in competence ratings between confidence levels is substantially larger for the male CEO (1.108) (p < .001) than the female CEO (0.717) (p < .001).

Table 5

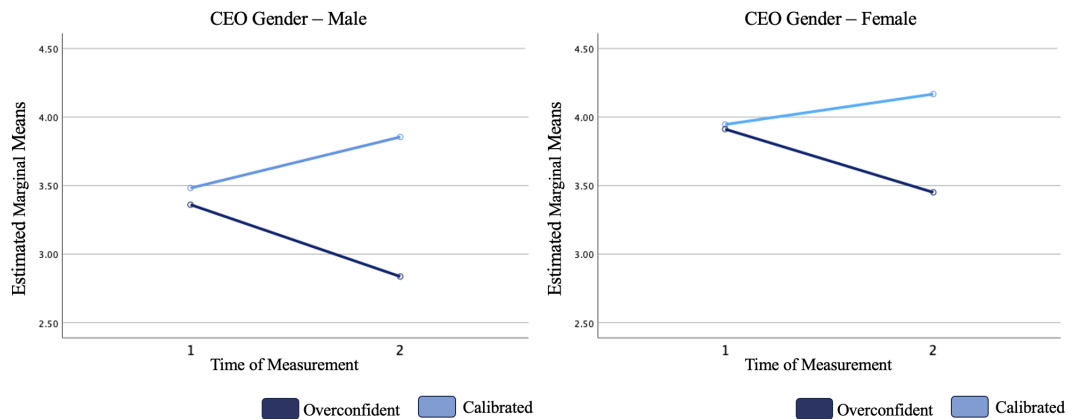*Competence Ratings – Phase 2*

Competence Ratings – Phase 2

Note. Competence ratings in Phase 2, depending on confidence level and CEO gender. Error bars: ± 1 SEM.

Changes From Phase 1 to Phase 2. To explore how evaluations of the CEOs changed when performance feedback was revealed, we conducted a 2 x 2 x 2 mixed ANOVA. The analysis demonstrated a main effect of time of measurement $F(1, 450) = 14.445$, $p < .001$, $\eta_p^2 = .031$, indicating that our manipulation had the intended effect. Summarizing across conditions, participants assigned lower scores in Phase 2 ($M = 3.58$, $SD = .87$) than in Phase 1 ($M = 3.68$, $SD = .78$). There was no interaction between time of measurement and CEO gender $F(1, 450) = .729$, $p = .394$, $\eta_p^2 = .002$. That is, both the male and female CEO was affected similarly when participants received objective performance feedback. As expected, there was an interaction between time of measurement and confidence level $F(1, 450) = 235.788$, $p < .001$, $\eta_p^2 = .344$. Across genders, the well-calibrated CEO was evaluated as more competent in Phase 2 ($M = 4.01$, $SD = .048$) than in Phase 1 ($M = 3.71$, $SD = .049$), supporting H2. In contrast, the overconfident CEO was evaluated as less competent in Phase 2 ($M = 3.14$, $SD = .048$) than in Phase 1 ($M = 3.64$, $SD = .048$), supporting H3. Interestingly, the 2 x 2 x 2 ANOVA revealed a significant three-way interaction effect between time of measurement, CEO gender, and confidence level $F(1, 450) = 4.330$, $p = .038$, $\eta_p^2 = .010$. Post-hoc tests (Bonferroni corrected) showed that the male CEO (0.373 points) ($p < .001$) was rewarded more than the female CEO (0.222 points) ($p <$.

001) when revealed to be well-calibrated. Contradictory to H6, the male CEO (-0.524 points) ($p < .001$) was penalized slightly harder when revealed to be overconfident than the female CEO (- 0.461 points) ($p < .001$).

Table 6
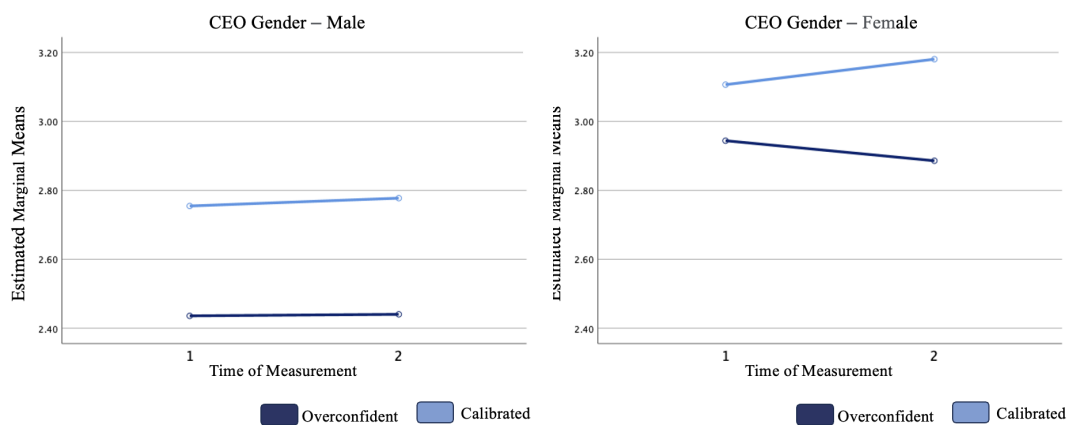
Competence Ratings Across Time of Measurement

Note. Competence ratings across confidence level, time of measurement, and CEO gender.

4.2 Warmth

We initially treated warmth as a secondary variable as we expected evaluations of warmth to remain largely unaffected (H5). Nevertheless, we choose to report some unexpected findings. In Phase 1, when participants did not have information to verify predictions, a 2 x 2 ANOVA demonstrated a significant main effect of confidence level on ratings of warmth $F(1, 450) = 9.575$, $p = .002$ $\eta_p^2 = .021$. Summarizing across conditions, the modest CEO ($M = 2.93$, $SD = .85$) was perceived as significantly warmer than the confident CEO ($M = 2.70$, $SD = .86$). This finding was rather surprising, as confidence level did not affect competence ratings before performance feedback, yet it did so with ratings of warmth. As with competence ratings, we found a significant main effect of CEO gender $F(1, 450) = 30.583$, $p < .001$, $\eta_p^2 = .064$. Participants assigned significantly higher warmth ratings for the female CEO ($M = 3.02$, $SD = .84$) than the male CEO ($M = 2.60$, $SD = .83$). Similar results were found for Phase 2 ratings.

We conducted a 2 x 2 x 2 mixed ANOVA to explore how evaluations of warmth changed when objective performance data was provided. As expected, we did not find a significant main effect of time of measurement $F(1, 450) = .554$, $p = .457$, $\eta_p^2 = .001$. However, we did find a significant interaction between time of measurement and confidence level $F(1, 450) = 6.817$, $p = .009$, $\eta_p^2 = .015$. Thus, failing to support H5, wherein we expected evaluations of warmth to remain largely unaffected when performance data was revealed. Interestingly, we unexpectedly found a significant three-way interaction between time of measurement, CEO gender, and confidence level $F(1, 450) = 3.902$, $p = .049$, $\eta_p^2 = .009$. To examine the nature of the interaction, we conducted pairwise comparisons (Bonferroni-corrected). Post-hoc tests showed that the significant interaction between time of measurement and confidence level was solely caused by the female CEO. In the well-calibrated condition, the female CEO was rewarded with slightly higher ratings of warmth ($p = .010$). Conversely, the female CEO was penalized in the overconfident condition ($p = .042$). No such effects were observed for the male CEO; Ratings of warmth were constant in both the calibrated ($p = .438$) and overconfident ($p = .878$) conditions after performance was revealed.

Table 7

Warmth Ratings Across Time of Measurement

Note. Warmth ratings across confidence level, time of measurement, and CEO gender

4.3 Additional Analyses

Several control variables (i.e., demographic variables and attitude scales) were measured in the survey. The majority of the demographic variables measured did not relate strongly to the dependent variables or did so in theoretically unimportant ways. Exceptions are as follows:

We found a main effect of participant gender on competence ratings at Phase 1, $F(1, 446) = 18.545$ $p < .001$ $\eta_p^2 = .040$, but no interaction effects, $p's$ ranging from .055 to .575. Summarizing across conditions, post-hoc tests (Bonferroni corrected) showed that female respondents generally assigned better competence ratings than male respondents, the only exception being for the well-calibrated male CEO ($p = .969$). The main effect of gender on competence ratings persisted at Phase 2, $F(1, 446) = 6.409$ $p = .012$, $\eta_p^2 = .014$. Interestingly, post-hoc tests (Bonferroni corrected) showed that female and male competence evaluations merged for all conditions ($p's$ ranging from .259 to .821) *except* when the female CEO was overconfident ($p = .001$), wherein male respondents rated the CEO as 0.431 points less competent. We find this interesting as there was no such difference when the male CEO was revealed as overconfident (0.069 points, $p = .607$).

Furthermore, the attitude scales included in the survey yielded some interesting findings. Agreeing that a leader should appear optimistic during interactions with employees, customers, and the media was correlated with positive evaluations of competence and warmth when the CEO was overconfident at both Phase 1 and Phase 2, $r's$ between .171 and .282, $p's$ between < .001 and .010. No such correlations were found when the CEO was well-calibrated, $r's$ between .033 and .086, $p's$ between .200 and .625. Agreeing that a leader should appear certain during interactions with employees, customers, and the media was positively related to competence and warmth ratings for the overconfident CEO ($r's$ between .232 and .292, all $p's$ < .001.) However, it was only positively correlated with competence for the well-calibrated CEO ($r's$ between .039 and .184, $p's$ between .006 and .564).

The other constructs we attempted to measure (i.e., age, education, leadership position, and company) did not relate strongly to the dependent variables and will not be discussed.

5 Discussion

5.1   Summary of Key Findings

All else equal, participants rated the well-calibrated CEO significantly more favorably than the overconfident CEO after actual performance was revealed, supporting H4. Even before performance feedback was provided, the modest CEO was evaluated somewhat better (but only in terms of warmth), failing to support H1. As hypothesized, the overconfident CEO faced social penalties (H3), while the well-calibrated CEO was rewarded upon discovery (H2). Thus, the present study clearly suggests that calibration trumps overconfidence. Furthermore, the female CEO was rated significantly more favorably than the male CEO across conditions. Yet, the female CEO was *not* penalized harder when confidence was revealed as unwarranted, failing to support H6. Contrary to H5, performance feedback did indeed influence warmth scores, but interestingly, this effect was observed only for the female CEO. None of the demographic variables measured had an impact (except for rater gender), supporting the robustness of our findings.

5.2   Interpersonal Consequences of Overconfidence

5.2.1   Modesty versus Confidence

Building on previous research, we hypothesized that the confident CEO would initially be perceived as more competent than the modest CEO. Contrary to H1, we did not obtain a statistically significant difference between the confident and modest CEO on competence ratings before performance feedback was provided. If anything, participants actually preferred the modest CEO, as (s)he was assigned significantly higher warmth ratings. Thus, our findings do not add to the large body of literature that indicates that the display of confidence increases interpersonal influence in beneficial ways.

Then, how can we understand these findings? Although the display of confidence has been found to predict perceptions of competence, common adjectives such as "arrogant," "know-it-all," "cocky," and "boastful" provide clear cues that people do not necessarily *like* confident individuals (Tenney et al., 2008). This notion is supported by Thoma (2016), who found that confident individuals are perceived as less likable than modest individuals. If participants

initially disliked the confident CEO, s(he) may arguably have been penalized with lower competence scores as well (e.g., Rosenberg et al., 1968). The fact that both the confident female and male CEO were assigned rather low warmth scores ($M = 2.94$ and 2.44, respectively) support this explanation. Another interpretation is that participants may actually have assumed that confidence was inflated. As previously outlined, the confidence heuristic holds that people use confidence as a cue to judge someone's competence, according to which people assume that the most confident individuals are most likely to be correct (Price & Stone, 2004). This implies that people do not trust others solely because they act confidently. Rather, they believe that confidence is grounded in something unless proven otherwise (i.e., the presumption of calibration hypotheses) (e.g., Price & Stone, 2004; Thomas & McFayden, 1995). According to Tenney et al. (2008), people will generalize from any useful evidence that calls into question the target's calibration. Sah et al. (2013) found that when information is easy to calibrate, people are more likely to claim "bulls" on cheap claims of confidence. The present study was set in the context of an acquisition, which by default, entails a great risk due to the required investment (Kenton, 2022). Consequently, confidence expressions may not have been taken at face value by participants and hence, may not have been used to infer competence. Notably, this explanation is consistent with the presumption of calibration hypothesis. An alternative explanation is that the differences between the experimental conditions (confident versus modest) simply were too subtle to obtain an effect on competence ratings.

A notion of caution is that a Norwegian sample was used. It is worth considering the influence of social codes in Norway, such as the Law of Jante (Janteloven). The Law of Jante can be summarized as "Do not think you are anything special. Do not think you are better than us." (SNL, 2020). In essence, behaviors associated with confidence and seeking attention may attract social sanctions (Beltagui & Schmidt, 2017). Importantly, a considerable amount of research on the interpersonal effects of confidence is carried out using American samples (e.g., Price & Stone, 2004; Sah et al., 2013; Tenney & Spellman, 2011). The United States exhibits an inclination towards self-promotion, as evidenced by its high ranking on the self-promotion scale (Twenge & Campbell, 2009). This taken into consideration, it is reasonable to assume that preference for confidence/modesty is subject to cross-cultural differences. Thus, cultural variations may influence the interpersonal consequences of displayed confidence.

5.2.2 Calibration versus Overconfidence

As the present study suggests that confidence does not yield social benefits even *before* performance feedback is reviewed, it is unlikely to do so after confidence is revealed as unwarranted. Consistent with H4, the present study suggests that calibration trumps overconfidence regardless of CEO gender. This taken into consideration, the present study supports the body of research that suggests that overconfidence has a cost upon discovery. This raises one critical question: when is overconfidence penalized, and when do interpersonal benefits of confidence persist after actual performance is revealed?

Verbal versus Nonverbal Displays of Confidence

In everyday life, people primarily signal overconfidence through subtler forms such as tone of voice, posture, and other nonverbal cues (Ridgeway & Diekema, 1989). For instance, overconfident individuals often speak first and with authority, hold a straight posture, and show little deference to others (Anderson et al., 2012). Several studies have captured overconfidence primarily in terms of nonverbal behaviors. For instance, Anderson and Kilduff (2009) and Anderson et al. (2012) used face-to-face groups, using naturally occurring confidence where overconfident individuals displayed behaviors used by others to infer competence. Ronay et al. (2019) conducted both field experiments where targets were assessed in-person and laboratory experiments using video excerpts. However, the sound was removed so that the participants' only source of information was the targets' nonverbal behavior. Interestingly, one key similarity between these studies is that they all found that overconfidence did not damage the reputation. Although this can arguably be attributed to the ease of which overconfidence can be recognized by participants, Ronay et al. (2019) manipulated targets' perceived ability in one of the studies by creating four fake resumes coupled with the video excerpts aimed to cue either high or low competence. The experiment revealed that regardless of ability, expressed confidence increased perceived leadership potential. Furthermore, Kennedy et al. (2013) obtained similar results with objective feedback on task performance, wherein trained actors expressed confidence through tone of voice and posture. Importantly, these studies were designed to mimic how individuals learn overconfidence in the real world. These studies provide evidence that nonverbal displays of overconfidence significantly impact participants' perceptions of targets. However, given the operationalization

issues surrounding nonverbal displays of overconfidence, they raise concern as a reliable measure of overconfidence (Tenney et al., 2019).

Due to the measurement difficulties of nonverbal displays of overconfidence, many studies have captured overconfidence through verbal expressions. Importantly, confidence expressions in studies with excerpts are exclusively verbal, as excerpts, by default, do not allow for nonverbal displays of overconfidence. For instance, Tenney et al. (2007) employed written trial summaries, Tenney et al. (2008) used written witness depositions, Tenney and Spellman (2011) used excerpts from applicant interviews, Sah et al. (2013) used photographs with advice in written form, and Vullioud et al., (2017) used messages from advisors. These studies all found that overconfident targets experienced a backlash when performance was revealed. The present study adds to the body of literature using verbal displays of overconfidence, utilizing audio clips with both precise numerical estimations and qualifying statements (e.g., "I am *convinced* this will strengthen our market share" versus "I *hope* this will strengthen our market share") (Brewer & Burke, 2002). This taken into consideration, an emerging pattern in the existing research is that studies featuring verbal displays of confidence typically find calibration to be beneficial, whereas studies featuring nonverbal displays of overconfidence find overconfidence to yield a net positive effect. The *plausible deniability hypothesis,* as proposed by Tenney et al. (2019), can shed light on this observation. According to this hypothesis, unwarranted confidence expressed verbally is easier to falsify than nonverbal overconfidence, which may yield stricter social penalties.

It is important to note that there are some exceptions to this pattern. For instance, studies by Paulhus (1998) and Anderson et al. (2008) both used laboratory groups where group members worked together, allowing for nonverbal expressions of confidence. In contrast to similar studies cited above, they found that overconfidence had a social cost upon discovery. However, both Paulhus (1998) and Anderson et al. (2008) measured overconfidence by comparing self-reports to observations made by others. Thus, overconfident individuals that have successfully spread their positive self-views will not appear overconfident on these measures, only those who appear unconvincing (Murphy et al., 2018). Furthermore, Price and Stone (2004) used two fictional financial advisors who predicted whether a stock would increase or decrease in value and presented stimuli with computers. Although the design eliminated all forms of nonverbal

confidence, they still found participants to prefer the overconfident advisor. However, Sah et al. (2011) suggest that this may simply be due to the ease of calibration. In Price and Stone's (2004) design, it required effort to calibrate the financial advisors. Additionally, the discrepancy between the modest and overconfident advisors was rather small, and participants may not have registered poor calibration (Sah et al., 2011). A notable observation is that the advisors were correct in the vast majority of trials (18 out of 24). Therefore, it could appear that the overconfident advisor was superior by assigning higher probabilities, e.g., 65 % versus 80 %. In the present study, participants were able to accurately infer overconfidence because the performance feedback visually displayed projections next to actual performance data. Additionally, control questions were included to ensure that all participants had registered the targets' overconfidence. This taken into consideration, claims were explicit and clearly falsifiable. This suggests that stating opinions with total certainty can yield a backlash from perceivers when opinions turn out to be wholly misguided.

Overconfidence among High-Level Leaders

The present study differs from the studies cited above by examining overconfidence specifically within the context of high-level leadership. This may impact the perception of overconfidence upon its discovery, as high-level leaders face different expectations than the general population (Schaumberg & Flynn, 2017). Interestingly, agreeing that a leader should appear optimistic/certain during interactions with stakeholders (i.e., employees, customers, and the media) was correlated with positive evaluations of the overconfident CEO. This suggests that some may regard small instances of overconfidence as forgivable sins of optimism (Armor et al., 2008). This notion is consistent with anecdotal evidence; overconfident leaders often maintain their influence in the workplace even when frequently making inaccurate claims (Tenney et al., 2019). One possible explanation is that displays of optimism/certainty are associated with charismatic leadership (i.e., leaders who are able to foster an impression that they and their mission are important) (Conger et al., 2000). Though not synonymous, overconfidence is arguably a prerequisite of charismatic leadership (Ronay et al., 2019). Thus, high-level leaders (particularly CEOs) may be excused for exaggerated confidence by personifying an inspiring aura of leadership charisma (Moore & Bazerman, 2022). Despite the conventional standard of unbiased predictions, there is indeed evidence that people sometimes prescribe optimism

over accuracy (Armor et al., 2008). For instance, some believe that an overconfident mindset, rather than a realistic mindset, is beneficial in situations that require effort and persistence (Armor et al., 2008). A recent meta-analysis by Burkhard et al. (2018) lends support to the notion that there may be merit to this belief. Contrary to common belief, they found a small ($r = .04$) yet significant relationship between CEO overconfidence and firm performance. This can, in part, be attributed to "a signaling value," where overconfident CEOs, *because* of their optimism, are more likely to encourage commitment among employees (Vitanova, 2021; Tenney et al., 2015). This taken into consideration, some individuals may grant high-level leaders some leeway when revealed as overconfident.

Although respondents that valued optimism and certainty perceived the overconfident CEO more favorably than the rest, s(he) was still, on average, penalized when exposed. While some may consider small errors forgivable, it is worth noting that the present study employed rather high levels of overconfidence. However, as corporate executives are often held accountable for firm performance (Farrell & Whidbee, 2003), the present study arguably mimics how overconfident individuals in high-level leadership positions are evaluated in the real world. Notably, most perceivers have limited, if any, social interaction with the leader. Perceivers usually become familiar with the leader through public appearances and occasional passes in the hallway. Thus, social interaction cannot counteract any negative publicity for making overconfident/and or inaccurate claims. Thus, objective performance indicators significantly impact the perceptions of high-level leaders (Rosette & Tost, 2010). This further suggests that firm performance should be influential, for example, consider an interesting observation made by Reyes et al. (2022). An extensive empirical analysis of 1,712 companies revealed that in expansion periods, firms with overconfident CEOs tend to outperform those with non-overconfident CEOs, despite more errors (Reyes et al., 2022). In contrast, overconfident CEOs tend to be harmful in recession periods due to greater risk-taking (Reyes et al., 2022). This further indicates that social penalties should be less pronounced when firm performance is good. Nevertheless, as we substantially overestimated performance, participants evaluated the overconfident CEOs in a "loss frame" despite the reasonably good outcome (Kahneman & Tversky, 1979).

Returning to the research question; "Is it more beneficial for leaders to be

well-calibrated than overconfident when actual abilities are revealed to others?" the present study clearly suggests so. However, it is important to keep in mind that we used a controversial test of our hypotheses. Overconfidence was expressed verbally, and claims were explicit and easily disprovable. However, as overconfidence is often expressed nonverbally in everyday life, it may be easier to get away with overconfidence than the current paper implies. Particularly as some may consider minor occurrences of overconfidence among leaders forgivable. Nevertheless, considering the role of firm performance in shaping perceptions of high-level leaders (Rosette & Tost, 2010), the present study arguably mirrors (to some extent) the way they are evaluated in real-world situations.

5.3   The Moderating Effect of CEO Gender

In the present study, we also sought to answer the following research question: "Are female leaders penalized harder when revealed as overconfident than equally overconfident male leaders?" Interestingly, our findings do not support conventional theories of gender and leadership that suggest that females are punished more harshly for failure and suffer greater reputational losses than otherwise similar men (e.g., Brescoll et al., 2010; Eagly & Carli, 2007; Fisk & Overton, 2019; Kennedy et al., 2016; Sarsons, 2017). The female CEO did not receive harsher punishment than the male CEO in terms of competence ratings. On the contrary, the male CEO was penalized slightly harder, though the difference (0.063 points) was marginal. This taken into consideration, why did we fail to obtain an effect? Though mere speculations, we propose several potential explanations as to why our findings contradict the research cited above.

5.3.1   The Influence of Contextual Factors

First and foremost, the present study employed a sample from Norway, a country widely recognized for its high degree of gender equality (World Economic Forum, 2022). It is important to recognize that this distinctive attribute of the sample is likely to exert an impact on the findings. Furthermore, an observation made by Tinsley et al. (2009) may yield some insights. Tinsley et al. (2009) propose that backlash effects are a context-dependent phenomenon. In particular, they suggest that female leaders suffer little or no backlash when gender stereotypes are not activated. In the present research, there were no aspects that should elicit gender stereotyping. As experimental conditions (CEO gender)

were kept as identical as possible, the design was likely to attenuate the likelihood of a backlash rather than amplify it. That is, the manipulation was arguably too weak to activate gender stereotypes. An alternative, though not contradictory, explanation stems from research by Amanatullah and Tinley (2008). They explored whether backlash was reduced depending on hierarchical position and found that backlash was diminished in senior leadership positions. To explain this finding, they proposed that females who occupy high-status positions are examined through a "status- and position lens" rather than a "gender lens" (Tinsley et al., 2009). If participants in the present study evaluated the female CEO through a "status- and position lens," the female CEO would be unlikely to challenge gender stereotypes and, therefore, be unlikely to receive harsher social penalties than the male CEO when revealed as overconfident. Consistent with this, Shcaumberg and Flynn (2017) suggest that hierarchical rank can moderate backlash effects against female leaders as occupying a high-status leadership role (e.g., being an elected CEO) dictates a level of social approval.

5.3.2   Female Leadership Advantage

An alternative interpretation stems from an emerging debate in the academic literature about the potential existence of a female advantage. There is evidence that older studies find male leaders to be perceived as more effective, whereas newer studies favor female leaders (Elsesser & Lever, 2011; Koenig et al., 2011). This can arguably be attributed to the emerging focus on empowering and collaborative leadership styles, as the view of leadership is taking a more "feminine" outlook (Koenig et al., 2011), decreasing the perceived incongruity between the requirements of leadership roles and characteristics of females.

Double Standards of Competence Model

In an extensive meta-analysis ($K = 99$), Paustian-Underdahl et al. (2014) found that female leaders were actually rated as more effective than males when they held senior-level management positions. This finding is consistent with the present research, as the female CEO was evaluated significantly more favorably than the male CEO across conditions. Yet, they run contrary to role congruity theory as stereotypes are likely to be most masculine for high-status, senior leadership positions, increasing role incongruity for females in these positions (Eagly & Karau, 2002). A possible explanation comes from the *double standards of competence model,* which proposes that females' mere presence in the highest

positions provides evidence of their abilities (Foschi, 2000). That is, people assume she must have been exceptionally competent to have reached such a top position despite the barriers to female career advancement. Recall that people appear to be aware of this gendered discrimination (Fisk & Overton, 2019). A laboratory study by Rosette and Toast (2010) provides support for this explanation, as they found that high-ranking female leaders received more positive evaluations than their male counterparts because they were thought to have faced higher standards. This can shed light on our findings, as participants had limited information and may have assumed that the female CEO, by default, must possess particularly high levels of competence to have reached a top position. Consequently, participants may not necessarily have attributed the inflated performance estimates to low ability (Swim & Sanna, 1996), at least not any more than for the male CEO. The double standards of competence model may also pose a possible explanation to why the male CEO was rewarded with higher competence ratings than the female CEO when revealed as well-calibrated, as participants may have needed more evidence about his abilities to determine competence. That is, as the male CEO had a worse starting point, he may have been rewarded more than the female CEO when ability was "confirmed."

However, in our examination of rater gender, we observed a tendency for male raters to exhibit a stronger gender-role congruity bias than female raters. Interestingly, male raters assigned significantly lower competence scores to the overconfident female CEO than did female raters. Yet, there was no such difference for the overconfident male CEO. This finding is consistent with a considerable body of research that suggests that compared to females, men are more likely to hold traditional stereotypes about females (e.g., Koenig et al., 2011; Schein, 2001; Koch et al., 2015).

Expectancy Violation Theory

A recent study by Ma et al. (2022) offers an alternative and somewhat contradictory explanation for the existence of a female advantage. They focused on the display of agentic traits and emphasized that some agentic content appears positively valenced (e.g., competent, diligent), whereas others appear negatively valenced (e.g., aggressive, dominant) (Ma et al., 2022). They argued that positively valenced agentic traits (prescriptions) should be socially desirable to possess regardless of gender, and negative valenced agentic traits (proscriptions), should not be desirable to possess for either gender (Ma et al., 2022). According

to a classic article by Prentice and Carranza (2002), agency prescriptions (e.g., competent, diligent) are strong prescriptions for men and weak prescriptions for females. In contrast, agency proscriptions (e.g., aggression, dominance) are weak proscriptions (i.e., allowable transgressions) for men and strong proscriptions for females. That is, they vary in terms of how strongly they are imposed on males and females. Drawing on *expectancy violation theory* (Jussim et al., 1987), Ma et al. (2022) found that an agentic advantage was elicited when female leaders violated agency prescriptions, as they displayed more favorable qualities than expected. Conversely, an agentic disadvantage was elicited when they violated agency proscriptions, as they displayed more undesirable qualities than expected (Ma et al., 2022).

Interestingly, the article by Ma et al. (2022) can shed light on some of our more unexpected findings. When revealed as well-calibrated, the CEO displayed favorable agentic qualities by appearing competent. Conversely, when revealed as overconfident, the CEO displayed negative agentic qualities by coming across as self-promoting/arrogant. Expectancy violation theory suggests that the female CEO should be rewarded more than the male CEO in the well-calibrated condition and punished more than the male CEO in the overconfident condition. Although we did not obtain any such effects on competence ratings, we did so with warmth ratings. Consistent with expectancy violation theory, the female CEO was rewarded with significantly higher ratings of warmth in the well-calibrated condition and penalized with significantly lower warmth ratings in the overconfident conditions. This is arguably driven by an expectancy violation as the female CEO was perceived to possess more favorable/unfavorable qualities than expected (due to gender stereotypes), eliciting more "extreme" interpersonal evaluations in the direction of the expectancy violation (though the effect size was rather small) (Ma et al., 2022). Conversely, as communal traits merely constitute weak prescriptions for males, warmth ratings remained unaffected for the male CEO when actual performance was revealed. This taken into consideration, the present research suggests that perceived likability is somewhat more tenuous and fragile for female than male leaders (e.g., Williams and Dempsey, 2014). Building on research by Prentice and Carranza (2002), this may simply be because participants are more attuned to communal traits for female leaders, as people strongly expect females to possess these traits. Noteworthy, this explanation is consistent with the double bind, which proposes that female leaders are expected

to be competent and warm (Trzebiatowski et al., 2023). Overall, this suggests that by displaying desirable agentic qualities, female leaders not only avoid social penalties but can gain an advantage in leadership evaluations relative to their male counterparts. Conversely, female leaders stand more to lose by displaying non-desirable agentic qualities (Ma et al., 2022).

Returning to our research question, the present study adds to the body of research that suggests a female advantage as the female CEO was initially evaluated more favorably yet was not penalized with lower competence ratings than the male CEO when revealed as overconfident. These are promising findings as they indicate that stereotypes of gender and leadership may be shifting (Diekman & Eagly, 2000).

5.4   Limitations, Strengths, and Future Research

As with all studies, we acknowledge the existence of some limitations. An experimental design was employed, with random assignment to the experimental conditions. The experimental design incorporates a high degree of control over extraneous variables, increasing internal validity. This allows us to establish causal relationships between variables (Bordens & Abbott, 2014, p. 105). However, some factors that may threaten internal validity must be addressed. Although the experimental conditions were kept as identical as possible, some factors were difficult to control for with the current study design. For instance, different actors were used for the male- and female CEO. Thus, we cannot rule out the possibility that some participants may have preferred the appearance or voice of the female CEO (e.g., Hosoda et al., 2003; Rosenberg et al., 1968).

On the flip side, high internal validity may potentially limit the ecological validity of the findings (Bordens & Abbott, 2014, p. 115). That is, the findings may not necessarily replicate in real-world settings as the experiment utilized an artificial scenario (i.e., fictitious podcast). For instance, the present study differs from that of the real world as overconfident leaders often have the opportunity to modify their behavior to appease others to maintain a positive standing (Kennedy et al., 2013). Additionally, performance feedback is seldom as objective and explicit in most organizations. Thus, in the real world, detecting overconfidence and seeing its cost may be more difficult. However, we aligned the survey experiment as closely as possible with characteristics of that of an actual acquisition, incorporating objective performance measures present in real-world

acquisitions. Additionally, the research design is highly consistent with the type of perceptions made by the general public about prominent leaders (Rosette & Tost, 2010).

Furthermore, a large sample size ($N=454$) was obtained. However, excluding participants who had substantial missing data or failed to respond correctly to one or more control questions may introduce limitations to our study. This exclusion may introduce sampling bias, as the characteristics and responses of the excluded participants may differ systematically from those who completed the survey (Bordens & Abbott, 2014, p. 161). Moreover, as the study was exclusively conducted with a Norwegian sample, there arises a concern regarding the generalizability of the findings beyond the sample (Bordens & Abbott, 2014, p. 159). Preferences for confidence or modesty may be contingent upon cultural variations. Despite cross-cultural consistencies regarding the content of gender stereotypes (Best & Williams, 2001), it is imperative to recognize the variability in the extent to which gender norms are enforced globally. This taken into consideration, the interpersonal consequences of overconfidence and the extent to which gender moderates this relationship should be scrutinized in a variety of cultures. Despite this caveat, we used samples from two substantially different Norwegian companies. Veidekke is a male-dominated entrepreneurial company with a strikingly low percentage of female executives (10.1 %) (Veidekke, 2022). Gjensidige, on the other hand, is an insurance company with a relatively high percentage of female executives (55 %) (Modular Fiance 2023). We did not obtain any statistically significant differences between the two companies, which lends support to the robustness and reliability of our findings.

A more general concern is regarding inconsistencies in how research has captured overconfidence. Although the three types of overconfidence (i.e., overprecision, overestimation, and overplacement) are conceptually and empirically distinct, they have often been used interchangeably (Moore & Schatz, 2017). Furthermore, given the lack of verifiability and replicability for nonverbal displays of overconfidence, they constitute a questionable measure of overconfidence. Yet, they clearly shape participants' perceptions of targets (Tenney et al., 2019). Consequently, further research is warranted to explore more reliable measures of nonverbal overconfidence. This taken into consideration, it is difficult to determine whether results vary because of measurement idiosyncrasies of specific studies or whether they truly differ by account (Moore & Healy, 2008).

It is worth mentioning that the effect sizes in the present study are fairly small. Some critics have argued that such small effects lack substantial importance (e.g., Vecchio, 2002). Nevertheless, numerous researchers have expressed disagreement regarding the effect size, arguing that even small effect sizes can hold practical importance in real-life settings (e.g., Abelson, 1985; Bushman & Anderson, 2001). For instance, small biases against female leaders can produce large consequences in terms of their ability to succeed in top leadership positions.

Moreover, we employed rather high levels of overconfidence. Future investigations should explore if the social penalties persist with lower levels of leader overconfidence. Additionally, it is important to note that performance was actually quite well in the present study. Although the well-calibrated CEO was rewarded in the present study, it is worth questioning whether s(he) would have the same advantage if actual performance was worse (or slightly worse than expected). Thus, future research should explore whether it is beneficial to be overconfident or well-calibrated when the modest target, too, fails to reach projections. Finally, the present study captured a one-off occurrence of overconfidence. To gain further insights, future research should examine whether the drawbacks associated with overconfidence and gender differences persist when individuals encounter such occurrences repeatedly over time.

5.4   Implications

With these caveats in mind, the present study has several implications on an individual, organizational, and societal level. On the individual level, it has implications for impression management, an essential aspect of corporate leadership (Pfeffer, 1981). Impression management involves strategic behavior aimed at establishing perceptions of competence and likability while avoiding unfavorable impressions (Bourdage et al., 2018). The present study suggests that high-level leaders should strategically construct their public image as well-calibrated individuals, striving for accuracy in their self-assessment and avoiding exaggerated confidence, particularly explicit and falsifiable claims. This approach aligns with societal norms in Norway, such as Law of Jante, and can enhance the perception of competence and likability. On the organizational level, the present study suggests that firms may benefit from considering modesty and calibration as desirable traits when selecting and evaluating leaders. Identifying individuals who exhibit accurate self-assessment and avoid overconfidence can

arguably lead to more positive evaluations from stakeholders, such as employees, investors, customers, and the media. This applies regardless of whether the leader is male or female. On the societal level, the observation that the female CEO was evaluated significantly more favorably than the male CEO indicates a potential reduction in gender biases. This finding strengthens the idea of advocating for a more equitable society and indicates that ongoing efforts should be made to boost the representation of females in prominent leadership positions. Note, however, that the findings of the present study are primarily applicable to large corporations in Norway, given that the sample was drawn from this specific population.

6 Conclusion

The present study clearly suggests that it is more beneficial to be well-calibrated than overconfident in the realm of leadership perception. The superiority of calibration held true even *before* confidence was revealed as unwarranted, indicating a societal inclination towards valuing modesty (consistent with Law of Jante). This further suggests that preference for confidence versus modesty may be contingent upon cultural variations. The present study adds to the existing literature on overconfidence and suggests that corporate leaders should exercise caution when making claims that can be objectively refuted as explicit and falsifiable claims can elicit a backlash when proven erroneous. Future research should delve deeper into the interpersonal consequences of nonverbal displays of overconfidence, perhaps by addressing the challenges associated with operationalizing such behaviors. In contrast to conventional theories on gender and leadership, the present study indicates that female leaders are *not* penalized harder when revealed as overconfident than equally overconfident male leaders. On the contrary, as the female was evaluated significantly more favorably across conditions, the present study lends support to the potential existence of a female advantage. This provides encouraging evidence of changing attitudes toward female leaders and constitutes exciting avenues for future research.

7 References

Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin, 97*(1), 129–133. https://doi.org/10.1037/0033-2909.97.1.129

Albanesi, S., Olivetti, C., & Prados, M. J. (2015). *Gender and dynamic agency: Theory and evidence on the compensation of top executives*. (IDEAS working paper No. 718). http://hdl.handle.net/10419/120820

Amanatullah, E. T., & Tinsley, C. H. (2013). Punishing female negotiators for asserting too much…or not enough: Exploring why advocacy moderates backlash against assertive female negotiators. *Organizational Behavior and Human Decision Processes, 120*(1), 110–122. https://doi.org/10.1016/j.obhdp.2012.03.006

Anderson, C., Ames, D. R., & Gosling, S. D. (2008). Punishing hubris: The perils of overestimating one's status in a group. *Personality and Social Psychology Bulletin, 34*(1), 90–101. https://doi.org/10.1177/0146167207307489

Anderson, C., & Kilduff, G. J. (2009). Why do dominant personalities attain influence in face-to-face groups? The competence-signaling effects of trait dominance. *Journal of Personality and Social Psychology, 96*(2), 491–503. https://doi.org/10.1037/a0014201

Anderson, C., Brion, S., Moore, D. A., & Kennedy, J. A. (2012). A status-enhancement account of overconfidence. *Journal of Personality and Social Psychology*, *103*(4), 718–735. https://doi.org/10.1037/a0029395

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science, 28*(11), 1547–1562. https://doi-org.ezproxy.library.bi.no/10.1177/0956797617723724

Arkes, H. R., Christensen, C., Lai, C., & Blumer, C. (1987). Two methods of reducing overconfidence. *Organizational Behavior and Human Decision Processes*, *39*(1), 133–144. https://doi.org/10.1016/0749-5978(87)90049-5

Armor, D. A., Massey, C., & Sackett, A. M. (2008). Prescribed optimism: Is it right to be wrong about the future? *Psychological Science, 19*(4), 329–331. https://doi.org/10.1111/j.1467-9280.2008.02089.x

Barber, B. M., & Odean, T. (2016). Boys will be boys: gender,
overconfidence, and common stock investment. *The Quarterly Journal
of Economics, 116*(1), 261-292. https://doi.org/10.1162/003355301556400

Bell, E., Bryman, A., & Harley, B. (2019). *Business research methods* (5th ed.).
Oxford University Press.

Beltagui, A., & Schmidt, T. (2017). Why can't we all get along? A study of
hygge and janteloven in a danish social-casual games community.
Games and Culture, 12(5), 403–425.
https://doi.org/10.1177/1555412015590062

Best, D. L., & Williams, J. E. (2001). Gender and culture. In D. Matsumoto (Ed.),
The Handbook of Culture and Psychology, 195–219.
Oxford University Press.

Bhandari, G., & Deaves, R. (2006) The demographics of overconfidence. *The
Journal of Behavioral Finance, 7*(1), 5-11.
https://doi.org/10.1207/s15427579jpfm0701_2

Borden, K. S., & Abbott, B. B. (2014). *Research design and methods: A process
approach* (9th ed.). McGraw-Hill.

Bourdage, J. S., Roulin, N., & Tarraf, R. (2018). "I (might be) just that good":
Honest and deceptive impression management in employment interviews.
Personnel Psychology, 71(4), 597–632. https://doi.org/10.1111/peps.12285

Brescoll, V. L., Dawson, E., & Uhlmann, E. L. (2010). Hard won and easily lost:
The fragile status of leaders in gender-stereotype-incongruent
occupations. *Psychological Science, 21*(11), 1640–1642.
https://doi.org/10.1177/0956797610384744

Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and
eyewitness confidence on mock-juror judgments. *Law and Human
Behavior, 26*(3), 353–364. https://doi.org/10.1023/A:1015380522722

Budworth, M., & Mann, S. L. (2010). Becoming a leader: The challenge of
modesty for women. *Journal of Management Development*, *29*(2),
177–186. https://doi.org/10.1108/02621711011019314

Burgess, D., & Borgida, E. (1999). Who women are, who women should be:
Descriptive and prescriptive gender stereotyping in sex discrimination.
Psychology, Public Policy, and Law, 5(3), 665– 692.
https://doi.org/10.1037/1076-8971.5.3.665

Bushman, B. J., & Anderson, C. A. (2001). Media violence and the American

public: Scientific facts versus media misinformation. *The American Psychologist, 56*(6-7), 477–489. https://doi.org/10.1037/0003-066X.56.6-7.477

Burkhard, B., Siren, C. A., Van Essen, M., & Grichnik, D. (2018). A meta-analytic approach to understanding the effects of CEO overconfidence on firm performance. *Academy of Management Annual Meeting Proceedings, 1*. https:// doi.org/10.5465/AMBPP.2018.91

Catalyst. (2007). *The double-bind dilemma for women in leadership: Damned if you do, doomed if you don't.* Catalyst. https://www.catalyst.org/wp-content/uploads/2019/01/The_Double_Bind_Damned_if_You_Do_Doomed_if_You_Dont.pdf

Camerer, C. F., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review, 89*(1), 306–318. https://doi.org/10.1257/aer.89.1.306

Casler, M. D. (2015). Fundamentals of experiment design: Guidelines for designing successful experiments. *Agronomy Journal, 107*(1), 692–705. https://doi.org/10.2134/agronj2013.0114

Cejka, M. A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin, 25*(4), 413–423. https://doi.org/10.1177/0146167299025004002

Chen, S.-S., Ho, K.-Y., & Ho, P.-H. (2014). CEO overconfidence and long-term performance following R&D increases. *Financial Management, 43*(2), 245–269. https://doi.org/10.1111/fima.12035

Cheng, J. T., Anderson, C., Tenney, E. R., Brion, S., Moore, D. A., & Logg, J. M. (2021). The social transmission of overconfidence. *Journal of Experimental Psychology: General, 150*(1), 157–186. https://doi.org/10.1037/xge0000787

Conger, J. A., Kanungo, R. N., & Menon, S. T. (2000). Charismatic leadership and follower effects. *Journal of Organizational Behavior, 21*(7), 747-767. https://doi.org/10.1002/1099-1379(200011)21:7<747::AID-JOB46>3.0.CO;2-J

Core. (2022) *Topplederbaromoteret 200* [CEO Barometer 200]. https://samfunnsforskning.no/core/publikasjoner/core-indikator-status/core-topplederbarometer/2022/

Cramer, R. J., DeCoster, J., Harris, P. B., Fletcher, L. M., & Brodsky, S. L. (2011). A confidence-credibility model of expert witness persuasion: Mediating effects and implications for trial consultation. *Consulting Psychology Journal: Practice and Research, 63*(2)*,* 129–137. http://dx.doi.org/10.1037/a0024591

Cuddy, Glick, P., & Beninger, A. (2011). The dynamics of warmth and competence judgments and their outcomes in organizations. *Research in Organizational Behavior, 31*, 73–98. https://doi.org/10.1016/j.riob.2011.10.004

Daubman, K. A., & Sigall, H. (1997). Gender differences in perceptions of how others are affected by self-disclosure of achievement. *Sex Roles, 37*(1-2), 73–89. https://doi.org/10.1023/A:1025640720363

Diekman, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: women and men of the past, present, and future. *Personality & Social Psychology Bulletin, 26*(10), 1171–1188. https://doi.org/10.1177/0146167200262001

Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. Lawrence Erlbaum

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*(3), 573–598. https://doi.org/10.1037/0033-295X.109.3.573

Eagly, A. H., & Carli, L. L. (2007)*.* Through the labyrinth: The truth about how women become leaders. *Gender in Management, 24*(1). https://doi.org/10.1108/gm.2009.05324aae.001

Elsesser, K. M., & Lever, J. (2011). Does gender bias against female leaders persist? Quantitative and qualitative data from a large-scale survey. *Human Relations, 64*(12), 1555–1578. https://doi.org/10.1177/0018726711424323

Emerson, R. W. (2022). ANOVA Assumptions. *Journal of Visual Impairment & Blindness, 116*(4), 585–586. https://doi.org/10.1177/0145482X221124187

Engelen, A., Neumann, C., & Schwens, C. (2015). "Of course I can": The effect of CEO overconfidence on entrepreneurially oriented firms. *Entrepreneurship Theory and Practice, 39*(5), 1137–1160. https://doi.org/10.1111/etap.12099

Farrell, K. A., & Whidbee, D. A. (2003). Impact of firm performance expectations

on CEO turnover and replacement decisions. *Journal of Accounting & Economics, 36*(1), 165–196. https://doi.org/10.1016/j.jacceco.2003.09.001

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3*(4), 552-564. https://doi.org/10.1037/0096-1523.3.4.552

Fisk, S. R., & Overton, J. (2019). Who wants to lead? Anticipated gender discrimination reduces women's leadership ambitions. *Social Psychology Quarterly, 82*(3), 319–332. https://doi.org/10.1177/0190272519863424

Fiske, S. T., & Stevens, L. E. (1993). What's so special about sex? Gender stereotyping and discrimination. In S. Oskamp & M. Costanzo (Eds.), *Gender issues in contemporary society*, 173–196. Sage Publications.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences, 11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. *Current Directions in Psychological Science: Journal of the American Psychological Society, 27*(2), 67–73. https://doi.org/10.1177/0963721417738825

Foschi, M. (2000). Double standards for competence: Theory and research. *Annual Review of Sociology, 26*(1), 21–42. https://doi.org/10.1146/annurev.soc.26.1.21

Forbes. (2023). *World′s Billionaire List.* https://www.forbes.com/billionaires/

Gjensidige. (2022). *Annual report 2022*. Gjensidige Forsikring.

Glick, P., Wilk, K., & Perreault, M. (1995). Images of occupations: Components of gender and status in occupational stereotypes. *Sex Roles, 32*(9-10), 565–582. https://doi.org/10.1007/BF01544212

Gupta, V. K., Han, S., Mortal, S. C., Silveri, S., & Turban, D. B. (2018). Do women CEOs face greater threat of shareholder activism compared to male CEOs? A role congruity perspective. *Journal of Applied Psychology, 103*(2), 228–236. https://doi.org/10.1037/apl0000269

Halkias, G., & Diamantopoulos, A. (2020). Universal dimensions of individuals' perception: Revisiting the operationalization of warmth and competence with a mixed-method approach. *International Journal of Research*

Marketing, 37, 714-736. https://doi.org/10.1016/j.ijresmar.2020.02.004

Harkness J., & Schoua-Glusberg A (1998). Questionnaires in translation. In Harkness J. (Ed.), *Cross-Cultural Survey Equivalence,* 87-126. ZUMA. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-49733-1

Hardman, D. (2009). *Judgment and Decision Making: Psychological Perspectives* (1st ed.). John Wiley and Sons Ltd.

Hayward, M. L., & Hambrick, D. C. (1997). Explaining the premiums paid for large acquisitions: Evidence of CEO hubris. *Administrative Science Quarterly, 42*(1), 103–127. https://doi.org/10.2307/2393810

Hayward, M. L. A., Shepherd, D. A., & Griffin, D. (2006). A hubris theory of entrepreneurship. *Management Science, 52*(2), 160–172. https://doi.org/10.1287/mnsc.1050.0483

Heaton, J. B. (2002). Managerial optimism and corporate finance. *Financial Management, 31*(2), 33–45. https://doi.org/10.2307/3666221

Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. *Research in Organizational Behavior, 5*, 269–298.

Heilman, M. E. (1995). Sex stereotypes and their effects in the workplace: What we know and what we don't know. *Journal of Social Behavior and Personality, 10*(6), 3–26.

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues, 57*(4), 657–674.https://doi.org/10.1111/0022-4537.00234

Heilman, M. E., & Caleo, S. (2018). Combatting gender discrimination: A lack of fit framework. *Group Processes & Intergroup Relations, 21*(5), 725–744. https://doi.org/10.1177/1368430218761587

Ho, P.-H., Huang, C.-W., Lin, C.-Y., & Yen, J.-F. (2016). CEO overconfidence and financial crisis: Evidence from bank lending and leverage. *Journal of Financial Economics, 120*(1), 194–209. https://doi.org/10.1016/j.jfineco.2015.04.007

Hosoda, M., Stone-Romero, E. F., & Coats, G. (2003). The effects of physical attractiveness on job-related outcomes: A meta-analysis of experimental studies. *Personnel Psychology, 56*(2), 431–462. https://doi.org/10.1111/j.1744-6570.2003.tb00157.x

Huang, & Kisgen, D. J. (2013). Gender and corporate finance: Are male executives overconfident relative to female executives? *Journal of*

Financial Economics, 108(3), 822–839.

https://doi.org/10.1016/j.jfineco.2012.12.005

Johnson, D. D. P. (2004). *Overconfidence and war: The havoc and glory of positive illusions.* Harvard University Press. https://doi.org/10.2307/j.ctvk12rcg

Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology, 89*(6)*,* 899–913. https://doi.org/10.1037/0022-3514.89.6.899

Jussim, L., Coleman, L. M., & Lerch, L. (1987). The nature of stereotypes: A comparison and integration of three theories. *Journal of Personality and Social Psychology, 52*, 536–546. https://doi.org/10.1037/0022-3514.52.3.536

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An analysis of decision under risk. *Econometrica*, *47*(2). https://doi.org/10.2307/1914185

Kennedy, J. A., Anderson, C., & Moore, D. A. (2013). When overconfidence is revealed to others: Testing the status-enhancement theory of overconfidence. *Organizational Behavior and Human Decision Processes, 122,* 266-279. http://dx.doi.org/10.1016/j.obhdp.2013.08.005

Kennedy, J., McDonnell, M.-H., & Stephens, N. (2016). Does gender raise the ethical bar? Exploring the punishment of ethical violations at work. *Academy of Management Annual Meeting Proceedings, 2016*(1), 11664. https://doi.org/10.5465/ambpp.2016.11664abstract

Kenton, W. (2022). *What Is an Acquisition? Definition, Meaning, Types, and Examples*. Investopedia. https://www.investopedia.com/terms/a/acquisition.asp

Kim, H. S., & Jang, S. (2021). CEO Overconfidence and firm performance: The moderating effect of restaurant franchising. *Cornell Hospitality Quarterly, 62*(2), 276–292. https://doi.org/10.1177/1938965519899926

Koch, A. J., D'Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology, 100*(1), 128–161. https://doi.org/10.1037/a0036734

Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader

stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin, 137*(4), 616–642. https://doi.org/10.1037/a0023557

Kozak, M., & Piepho, H.‑P. (2018). What's normal anyway? Residual plots are more telling than significance tests when checking ANOVA assumptions. *Journal of Agronomy and Crop Science*, *204*(1), 86–98. https://doi.org/10.1111/jac.12220

Kervyn, N., Yzerbyt, V. Y., Judd, C. M., & Nunes, A. (2009). A question of compensation: The social life of the fundamental dimensions of social perception. *Journal of Personality and Social Psychology, 96*(4), 828–842. https://doi.org/10.1037/a0013320

Lee, J. M., Park, J. C., & Chen, G. (2023). A cognitive perspective on real options investment: CEO overconfidence. *Strategic Management Journal*, *44*(4), 1084-1110. https://doi.org/10.1002/smj.3469

Lee, J. M., Hwang, B.-H., & Chen, H. (2017). Are founder CEOs more overconfident than professional CEOs? Evidence from S&P 1500 companies. *Strategic Management Journal, 38*(3), 751–769. https://doi.org/10.1002/smj.2519

Li, J., & Tang, Y. (2010). CEO hubris and firm risk taking in China: The moderating role of managerial discretion. *Academy of Management Journal, 53*(1), 45–68. https://doi.org/10.5465/AMJ.2010.48036912

Lundeberg, M., Fox, P., & Punćcohaŕ, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology, 86*(1), 114–121. https://doi.org/10.1037/0022-0663.86.1.114

Ma, A., Rosette, A. S., & Koval, C. Z. (2022). Reconciling female agentic advantage and disadvantage with the CADDIS measure of agency. *Journal of Applied Psychology, 107*(12), 2115–2148. https://doi.org/10.1037/apl0000550

Malmendier, U., & Tate, G. (2005). CEO overconfidence and corporate investment. *The Journal of Finance, 60*(6), 2661–2700. http://www.jstor.org/stable/3694800

Malmendier, U., & Tate, G. (2005). Does CEO overconfidence affect corporate investment? CEO overconfidence measures revisited. *European Financial Management: The Journal of the European Financial*

Management Association 11(5), 649-659.
https://doi.org/10.1111/j.1354-7798.2005.00302.x

Malmendier, U., & Tate, G. (2008). Who makes acquisitions? CEO
overconfidence and the market's reaction. *Journal of Financial Economics,
89*(1), 20–43. https://doi.org/10.1016/j.jfineco.2007.07.002

Meikle, N. L., Tenney, E. R., & Moore, D. A. (2016). Overconfidence at work:
Does overconfidence survive the checks and balances of organizational
life? *Research in Organizational Behavior, 36*, 121–134.
https://doi.org/10.1016/j.riob.2016.11.005

Miller, D. J., Spengler, E. S., & Spengler, P. M. (2015). A meta-analysis of
confidence and judgment accuracy in clinical decision making. *Journal of
Counseling Psychology*, *62*(4), 553-567.
https://doi.org/10.1037/cou0000105

Mitchell, T. R., Thompson, L., Peterson, E., & Cronk, R. (1997). Temporal
adjustments in the evaluation of events: The 'rosy view.' *Journal of
Experimental Social Psychology, 33*(4), 421–448.
https://doi.org/10.1006/jesp.1997.1333

Modular Finance (2023). Endringer i Gjensidige-konsernets
organisasjonsstruktur og konsernledelse [Changes in Gjensidige's
organizational structure and management group].
https://mfn.se/ob/a/gjensidige-forsikrings/gjf-endringer-i-gjensidige-konse
rnets-organisasjonsstruktur-og-konsernledelse-lb09d8b0

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence.
Psychological Review, 115(2), 502–517.
https://doi.org/10.1037/0033-295X.115.2.502

Moore, D. A., & Schatz, D. (2017). The three faces of overconfidence. *Social and
Personality Psychology Compass*, *11*(8), e12331–n/a.
https://doi.org/10.1111/spc3.1233

Moore, D. A., & Bazerman, M. H. (2022). Leadership & overconfidence.
Behavioral Science & Policy, 8(2), 59–69.

Murphy, S. C., Barlow, F. K., & von Hippel, W. (2018). A longitudinal test of
three theories of overconfidence. *Social Psychological & Personality
Science, 9*(3), 353–363. https://doi.org/10.1177/1948550617699252

Nichols, A. L., & Cottrell, C. A. (2014). What do people desire in their leaders?
The role of leadership level on trait desirability. *The Leadership Quarterly,*

25(4), 711–729. https://doi.org/10.1016/j.leaqua.2014.04.001

Oh, D., Shafir, E. & Todorov, A. (2020). Economic status cues from clothes affect perceived competence from faces. *Nature Human Behavavior 4,* 287–293 https://doi.org/10.1038/s41562-019-0782-4

Paulhus, D. L. (1998). Interpersonal and intrapsychic adaptiveness of trait self-enhancement: A mixed blessing? *Journal of Personality and Social Psychology, 74*(5), 1197–1208. https://doi.org/10.1037/0022-3514.74.5.1197

Paustian-Underdahl, S. C., Walker, L. S., & Woehr, D. J. (2014). Gender and perceptions of leadership effectiveness: A meta-analysis of contextual moderators. *Journal of Applied Psychology, 99*(6), 1129–1145. https://doi.org/10.1037/a0036751

Pfeffer, J. (1981). Management as symbolic action: the creation and maintenance of organizational paradigm. *Research in organizational behavior, 3*, 1-52.

Phalet, K., & Poppe, E. (1997). Competence and morality dimensions of national and ethnic stereotypes: A study in six eastern-European countries. *European Journal of Social Psychology, 27*(6), 703–723. https://doi.org/10.1002/(SICI)1099-0992(197711/12)27:6<703::AID-EJSP 841>3.0.CO;2-K

Piepho, H. P., Möhring, J., & Williams, E. R. (2013). Why randomize agricultural experiments? *Journal of Agronomy and Crop Science, 199*(5), 374–383. https://doi.org/10.1111/jac.12026

Plous, S. (1993). *The psychology of judgment and decision making*. Mcgraw-Hill Book Company.

Prentice, D. A., & Carranza, E. (2002). What women and men should be, shouldn't be, are allowed to be, and don't have to be: The contents of prescriptive gender stereotypes. *Psychology of Women Quarterly, 26*(4), 269–281. https://doi.org/10.1111/1471-6402.t01-1-00066

Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *Journal of Behavioral Decision Making*, *17*(1), 39–57. https://doi.org/10.1002/bdm.460

Reyes, T., Vassolo, R. S., Kausel, E. E., Torres, D. P., & Zhang, S. (2022). Does overconfidence pay off when things go well? CEO overconfidence, firm performance, and the business cycle. *Strategic Organization, 20*(3), 510–540. https://doi.org/10.1177/1476127020930659

Ridgeway, C. L., & Diekema, D. (1989). Dominance and collective hierarchy formation in male and female task groups. *American Sociological Review, 54*(1), 79–93. https://doi.org/10.2307/2095663

Ronay, R., Oostrom, J. K., Lehmann-Willenbrock, N., Mayoral, S., & Rusch, H. (2019). Playing the trump card: Why we select overconfident leaders and why it matters. *The Leadership Quarterly*, *30*(6), 101316. https://doi.org/10.1016/j.leaqua.2019.101316

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology, 9*(4), 283–294. https://doi.org/10.1037/h0026086

Rosette, A. S., & Tost, L. P. (2010). Agentic women and communal leadership. *Journal of Applied Psychology, 95*(2), 221–235. https://doi.org/10.1037/a0018204

Rudman, L. A. (1998). Self-promotion as a risk factor for women: The costs and benefits of counter-stereotypical impression management. *Journal of Personality and Social Psychology, 74*, 629–645.

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues, 57*(4), 743–762. https://doi.org/10.1111/0022-4537.00239

Rudman, L. A., & Phelan, J. E. (2008). Backlash effects for disconfirming gender stereotypes in organizations. *Research in Organizational Behavior, 28*, 61–79. https://doi.org/10.1016/j.riob.2008.04.003

Ryan, M. K., & Haslam, S. A. (2005). The glass cliff: evidence that women are over-represented in precarious leadership positions. *British Journal of Management, 16*(2), 81–90. https://doi.org/10.1111/j.1467-8551.2005.00433.x

Sah, S., Moore, D. A., & MacCoun, R. J. (2013). Cheap talk and credibility: The consequences of confidence and accuracy on advisor credibility and persuasiveness. *Organizational Behavior and Human Decision Processes*, *121*(2), 246–255. https://doi.org/10.1016/j.obhdp.2013.02.001

Sarsons, H. (2017). Recognition for group work: gender differences in academia. *The American Economic Review, 107*(5), 141–145. https://doi.org/10.1257/aer.p20171126

Schein, V. E. (1973). The relationship between sex role stereotypes and requisite

management characteristics. *Journal of Psychology, 57*(2), 95–100. https://doi.org/10.1037/h0037128

Schein, V. E. (1975). Relationships between sex role stereotypes and requisite management characteristics among female managers. *Journal of Applied Psychology, 60*(3), 340–344. https://doi.org/10.1037/h0076637

Schaumberg, R. L., & Flynn, F. J. (2017). Self-reliance: A gender perspective on its relationship to communality and leadership evaluations. *Academy of Management Journal, 60*(5), 1859–1881. https://doi.org/10.5465/amj.2015.0018

Semmler, C., Brewer, N., & Douglass, A. B. (2012). Jurors believe eyewitnesses. In B. L. Cutler (Ed.), *Convictions of the innocent: Lessons from psychological research,* 185–209. American Psychological Association. http://dx.doi.org/10.1037/13085-009

Shepperd, J. A., Ouellette, J. A., & Fernandez, J. K. (1996). Abandoning unrealistic optimism: Performance estimates and the temporal proximity of self-relevant feedback. *Journal of Personality and Social Psychology, 70*(4), 844–855. http://dx.doi.org/10.1037/0022- 3514.70.4.844

Sniezek, J. A., & Van Swol, L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational Behavior and Human Decision Processes, 84*(2), 288–307. https://doi.org/10.1006/obhd.2000.2926

SNL (2020). *Janteloven* [Law of Jante]. https://snl.no/Janteloven

Svenson, O (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica, 1981, 47*(2), 143-48.

Swim, J. K., & Sanna, L. J. (1996). He's Skilled, she's lucky: A meta-analysis of observers' attributions for women's and men's successes and failures. *Personality and Social Psychology Bulletin, 22*(5), 507–519. https://doi.org/10.1177/0146167296225008

Terborg, J. R. (1977). Women in management: A research review. *Journal of Applied Psychology, 62*(6), 647–664. https://doi.org/10.1037/0021-9010.62.6.647

Tenney, E. R., MacCoun, R. J., Spellman, B. A., & Hastie, R. (2007). Calibration trumps confidence as a basis for witness credibility. *Psychological Science, 18*(1), 46–50. https://doi.org/10.1111/j.1467-9280.2007.01847.x

Tenney, E. R., Spellman, B. A., & MacCoun, R. J. (2008). The benefits of knowing what you know (and what you don't): How calibration affects

credibility. *Journal of Experimental Social Psychology*, *44*(5), 1368–1375. https://doi.org/10.1016/j.jesp.2008.04.006

Tenney, E. R., & Spellman, B. A. (2011). Complex social consequences of self-knowledge. *Social Psychological and Personality Science*, *2*(4), 343–350. https://doi.org/10.1177/1948550610390965

Tenney, E. R., Logg, J. M., & Moore, D. A. (2015). (Too) optimistic about optimism: The belief that optimism improves performance. *Journal of Personality and Social Psychology, 108*(3), 377–399. https://doi.org/10.1037/pspa0000018

Tenney, E. R., Meikle, N. L., Hunsaker, D., Moore, D. A., & Anderson, C. (2019). Is overconfidence a social liability? The effect of verbal versus nonverbal expressions of confidence. *Journal of Personality and Social Psychology, 116*(3), 396–415. https://doi.org/10.1037/pspi0000150

Thoma, C. (2016). Under- versus overconfidence: An experiment on how others perceive a biased self-assessment. *Experimental Economics*, *19*(1), 218–239. https://doi.org/10.1007/s10683-015-9435-2

Thomas, J. P., & McFadyen, R. G. (1995). The confidence heuristic: A game-theoretic analysis. *Journal of Economic Psychology, 16*(1), 97–113. https://doi.org/10.1016/0167-4870(94)00032-6

Tinsley, C. H., Cheldelin, S. I., Schneider, A. K., & Amanatullah, E. T. (2009). Women at the bargaining table: Pitfalls and prospects. *Negotiation Journal, 25*(2), 233–248. https://doi.org/10.1111/j.1571-9979.2009.00222.x

Trzebiatowski, T., McCluney, C., & Hernandez, M. (2023). Managing the double bind: Women directors' participation tactics in the gendered boardroom. *Organization Science, 34*(2), 801–827. https://doi.org/10.1287/orsc.2022.1599

Twenge, J. M., & Campbell, W. K. (2009). *The narcissism epidemic: Living in the age of entitlement.* Free Press.

Van Teijlingen, E. & Hundley, V. (2002). The importance of pilot studies. *Nursing Standard, 16*(40)*,* 33-36. 10.7748/ns2002.06.16.40.33.c3214.

Vecchio, R. P. (2002). Leadership and gender advantage. *The Leadership Quarterly, 13*(6), 643–671. https://doi.org/10.1016/S1048-9843(02)00156-X

Vial, A. C., Napier, J. L., & Brescoll, V. L. (2016). A bed of thorns: Female

leaders and the self-reinforcing cycle of illegitimacy. *The Leadership Quarterly, 27*(3), 400–414. https://doi.org/10.1016/j.leaqua.2015.12.004

Vitanova, I. (2021). Nurturing overconfidence: The relationship between leader power, overconfidence and firm performance. *The Leadership Quarterly, 32*(4), 101342. https://doi.org/10.1016/j.leaqua.2019.101342

Vullioud, C., Clément, F., Scott-Phillips, T., & Mercier, H. (2017). Confidence as an expression of commitment: why misplaced expressions of confidence backfire. *Evolution and Human Behavior, 38*(1), 9–17. https://doi.org/10.1016/j.evolhumbehav.2016.06.002

Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal behavior, 24*(4), 239-263. https://doi.org/10.1023/A:1006610329284

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39*(5), 806–820. https://doi.org/10.1037/0022-3514.39.5.806

Weng, D. H., & Yamakawa, Y. (2022). I believe I can fly: how target venture CEO overconfidence affects acquisition completion. *Small Business Economics*. https://doi.org/10.1007/s11187-022-00679-6

Whitley, B. E., & Greenberg, M. S. (1986). The role of eyewitness confidence in Juror perceptions of credibility. *Journal of Applied Social Psychology, 16*(5), 387-409. https://doi.org/10.1111/j.1559-1816.1986.tb01148.x

Williams, J. E., & Best, D. L. (1990). *Measuring sex stereotypes: A multinational study*. Sage Publication

Williams, J. C., & Dempsey, R. (2014). *What Works for Women at Work: Four Patterns Working Women Need to Know*. York University Press.

Wojciszke, B. (1997). Parallels between competence- versus morality-related traits and individualistic versus collectivistic values. *European Journal of Social Psychology, 27*(3), 245–256. https://doi.org/10.1002/(SICI)1099-0992(199705)27:3<245::AID-EJSP81

World Economic Forum. (2022). *Global Gender Gap Report 2022*. https://www.weforum.org/reports/global-gender-gap-report-2022/

Zarnoth, P., & Sniezek, J. A. (1997). The social influence of confidence in group decision making. *Journal of Experimental Social Psychology, 33*(4), 345-366. https://doi.org/10.1006/jesp.1997.1326

8 Appendices

Appendix A
Translation of Competence and Warmth Items

Original version (English)	Norwegian translation
Competence	Kompetanse
1. (Capable)	1. (Kapabel)
2. Competent	2. Kompetent
3. Efficient	3. Effektiv
4. Skillful	4. Dyktig
5. (Industrious)	5. (Industriell)
6. Intelligent	6. Intelligent
Warmth	Varme
1. Friendly	1. Vennlig
2. Kind	2. Sympatisk
3. (Likable)	3. (Likandes)
4. Nice	4. Hyggelig
5. Warm	5. Varm

Note: Items derived from Halkias & Diamantopoulos (2020). The adjectives in parentheses were removed from the scale/survey: Capable, Industrious, and Likable

Examples of Survey Questions to Measure Competence and Warmth

Original version (Norwegian)

BI Norwegian Business School

Basert på førsteinntrykk, i hvilken grad oppfatter du Tor Johansen (ny CEO) som

	1 – svært liten grad	2	3	4	5 – svært stor grad
Vennlig	○	○	○	○	○
Hyggelig	○	○	○	○	○
Dyktig	○	○	○	○	○
Effektiv	○	○	○	○	○
Intelligent	○	○	○	○	○
Kompetent	○	○	○	○	○
Sympatisk	○	○	○	○	○
Varm	○	○	○	○	○

→

Translate version (English)

BI Norwegian Business School

Based on first impressions, to what extent do you perceive Tor Johansen (new CEO) as

	1 – very small degree	2	3	4	5 – very large degree
Efficient	○	○	○	○	○
Nice	○	○	○	○	○
Friendly	○	○	○	○	○
Kind	○	○	○	○	○
Competent	○	○	○	○	○
Skillful	○	○	○	○	○
Warm	○	○	○	○	○
Intelligent	○	○	○	○	○

→

Appendix B

Leader Optimism and Certainty Items

Note. Items derived from Armor et al. (2008)**.** Original (Norwegian) and translated (English) versions of the items that captured leader optimism and certainty

Participant Information Sheet and Demographic Questions

Original version (Norwegian)

Velkommen til denne undersøkelsen! Nedenfor følger et standard samtykkeskjema.

Formålet med prosjektet
Formålet med prosjektet er å undersøke hvordan ledere blir oppfattet. Du vil først høre et kort videoklipp, for deretter å svare på noen spørsmål tilknyttet dette.

Hvem er ansvarlig for forskningsprosjektet?
Handelshøyskolen BI er ansvarlig for prosjektet.

Hvorfor blir du bedt om å delta?
Vi ønsker å få tilgang til meningene til et bredt utvalg mennesker. For å delta må du være over 18 år.

Hva innebærer deltakelse for deg?
Undersøkelsen vil ta 2 - 5 minutter. Du vil bli spurt om dine inntrykk av en CEO i forbindelse med et oppkjøp. Dine svar vil bli registrert elektronisk. Deltakelse er frivillig. Hvis du velger å delta, kan du trekke deg når som helst.

Ditt personvern – hvordan vil vi lagre og bruke dine personlige data?
Vi vil IKKE registrere noen form for personlig data i dette spørreskjemaet (ikke IP-adresse, mail, eller annen informasjon som gjør deg identifiserbar). Datafilen vil være fullstendig anonym. Denne datafilen vil kunne bli delt med andre som er involvert i forskningsprosjektet. Når prosjektet er ferdigstilt, vil den anonymiserte filen også kunne deles med det vitenskapelige miljøet generelt.

Hvor kan jeg finne ut mer?

Dersom du ønsker å få resultatene tilsendt på mail eller har noen spørsmål til studien, ta gjerne kontakt med:

- Aleksandra Groos Dyvik, e-post: aleksandradyvik@gmail.com
- Marie Louise Nordbye Haraldsen, e-post: marie-louise.haraldsen@outlook.com
- Erik Løhre, Førsteamanuensis - Institutt for ledelse og organisasjon, Handelshøyskolen BI

Ettersom vi benytter et lydklipp kan det være lurt å ha AirPods/Headset tilgjengelig

Samtykke

Ved å klikke på "JA, JEG GIR MITT SAMTYKKE" gir jeg samtykke til å delta i denne undersøkelsen, og til at mine anonyme data kan brukes til forskningsøyemed (inkludert at data kan deles med andre forskere).

○ Ja, jeg gir mitt samtykke

→

Demografiske spørsmål

Vennligst svar på de følgende spørsmålene

Hva er alderen din?

[   ]

Hvilket kjønn identifiserer du deg som?

○ Kvinne

○ Mann

○ Ikke-binær / Annet

Hva er ditt høyeste fullførte utdanningsnivå?

○ Mastergrad eller høyere

○ Bachelorgrad

○ Videregående utdanning

○ Grunnskole

○ Vil ikke oppgi

Er du i en lederstilling?

○ Ja

○ Nei

○ Vil ikke oppgi

Hvilket selskap er du ansatt i? (frivillig)

[   ]

→

BI | Norwegian Business School

Welcome to this survey! Below is a standard consent form.

The purpose of the project

The purpose of the project is to investigate how leaders are perceived. You will first watch a short video clip and then answer some questions related to it.

Who is responsible for the research project?

BI Norwegian Business School is responsible for the project.

Why are you being asked to participate?

We want to access the opinions of a broad range of people. To participate, you must be over 18 years old.

What does participation entail for you?

The survey will take 2-5 minutes. You will be asked about your impressions of a CEO in connection with an acquisition. Your answers will be recorded electronically. Participation is voluntary. If you choose to participate, you may withdraw at any time.

Your privacy - how will we store and use your personal data?

We will NOT record any form of personal data in this questionnaire (no IP address, email, or other information that identifies you). The data file will be completely anonymous. This data file may be shared with others involved in the research project. When the project is completed, the anonymized file may also be shared with the scientific community in general.

Where can I find out more?

If you wish to receive the results by email or have any questions about the study, please contact:

- Aleksandra Groos Dyvik, email: aleksandradyvik@gmail.com
- Marie Louise Nordbye Haraldsen, email: marie-louise.haraldsen@outlook.com
- Erik Løhre, Associate Professor - Department of Leadership and Organization, BI Norwegian Business School

Since we use an audio clip, it may be useful to have AirPods/Headset available

Consent

By clicking "YES, I CONSENT" I consent to participate in this survey, and for my anonymous data to be used for research purposes (including the sharing of data with other researchers).

◯ Yes, I give my consent

BI Norwegian
Business School

Demographic Questions

Please answer the following questions.

What is your age?

What gender do you identify as?

○ Female

○ Male

○ Non-binary / Other

What is your highest completed level of education?

○ Master degree or higher

○ Bachelor degree

○ High school education

○ Elementary school

○ Will not disclose

Are you in a leadership position?

○ Yes

○ No

○ Will not disclose

Which company are you employed by? (optional)

→

62

Appendix D

Distribution of Respondents In the Four Different Experimental Conditions

Confidence Level	Gender		Total
	Female CEO *N*	Male CEO *N*	Total *N*
Modest (Phase 1)	115	110	225
Confident (Phase 1)	116	113	229
Total	231	223	454
Well-calibrated (Phase 2)	115	110	225
Overconfident (Phase 2)	116	113	229
Total	231	223	454

Appendix E
CEO Manuscripts

Overconfident CEO	
Norwegian (original script)	English (translated script)
Intervjuer: Hei, og velkommen til podkasten Ledelse- og Strategi. I dag har vi med oss Tor Johansen/Berit Johansen, ny CEO i Careto, en av Norges største forretningsbanker. Tor/Berit, bare noen få uker etter du tok over som CEO har du annonsert oppkjøpet av Ybank, en heldigital nettbank. Dette fremstår jo som et svært dristig valg. Hva ønsker du å oppnå ved dette?	**Interviewer:** Hi, and welcome to the Leadership and Strategy podcast. Today we are joined by Tor Johansen/Berit Johansen, newly appointed CEO of Careto, one of Norway's largest commercial banks. Tor/Berit, just a few weeks after assuming your role as CEO, you announced the acquisition of Ybank, a fully digital online bank. This appears to be a very bold move. What are your objectives with this acquisition?
Overkonfident Tor/Berit: Først og fremst skal vi ekspandere. Oppkjøpet av Ybank er en måte å oppnå vekst på kort tid. Jeg er sikker på at dette vil styrke vår markedsandel og gi oss et konkurransefortrinn. Særlig fordi oppkjøpet av Ybank gir oss tilgang til ny teknologi som garantert vil øke våre inntekter.	**Overconfident Tor/Berit:** First and foremost, we are going to expand. The acquisition of Ybank presents a way to attain rapid growth. I am convinced that this will strengten our market share and give us a competitive advantage. Particularly since acquiring Ybank will grants us access to new technology that undoubtedly will increase our revenue.

Intervjuer: Hvordan forventer du at oppkjøpet kommer til å påvirke nøkkeltallene til Careto med et 2-års perspektiv?	Interviewer: how do you expect the acquisition to affect Careto's key figures with a 2-year perspective?
Overkonfident Tor/Berit: Jeg forventer at vi når en markedsandel på privatmarkedet på 30 %, opp fra dagens nivå på 20 %. Altså en økning med 10% fra dagens andel. I tillegg forventer jeg at utlånsinntektene fra privatmarkedet vil øke fra dagens nivå på 10 milliarder til 15 milliarder. Altså en utlånsvekst på 50 %. Dette er selvsagt ambisiøst, men jeg er overbevist om at vi skal få det til.	**Overconfident Tor/Berit:** I expect us to reach a 30 % market share in the private market, up from our present 20 %, representing a 10 % increase. Furthermore, I expect the lending revenue from the private market to increase from its present level of 10 billion to 15 billion. In other words, a lending growth of 50 %. This is an ambitious target, but I have certan that we will achieve it.

Well-calibrated CEO	
Norwegian (original script)	English (translated script)
Intervjuer: Hei, og velkommen til podkasten Ledelse- og Strategi. I dag har vi med oss Tor Johansen/Berit Johansen, ny CEO i Careto, en av Norges største forretningsbanker. Tor/Berit, bare noen få uker etter du tok over som CEO har du annonsert oppkjøpet av Ybank, en heldigital nettbank. Dette fremstår jo som et	**Interviewer**: Hi, and welcome to the Leadership and Strategy podcast. Today we are joined by Tor Johansen/Berit Johansen, newly appointed CEO of Careto, one of Norway's largest commercial banks. Tor/Berit, just a few weeks after assuming your role as CEO, you announced the acquisition of Ybank, a fully digital online bank. This appears

svært dristig valg. Hva ønsker du å oppnå ved dette?	to be a very bold move. What are your objectives with this acquisition?
Velkalibrert Tor/Berit: Først og fremst ønsker vi å ekspandere. Oppkjøpet av Ybank er forhåpentligvis en måte å oppnå vekst på kort tid. Jeg håper at dette vil styrke vår markedsandel og gi oss et konkurransefortrinn. Særlig fordi oppkjøpet av Ybank gir oss tilgang til ny teknologi som potensielt kan øke våre inntekter.	**Well-calibrated Tor/Berit:** First and foremost, we want to expand. The acquisition of Ybank is hopefully a way to achieve growth in a short time. I hope that this will strengthen our market share and give us a competitive advantage. Especially because the acquisition of Ybank gives us access to new technology that can potentially increase our income.
Intervjuer: hvordan forventer du at oppkjøpet kommer til å påvirke nøkkeltallene til Careto med et 2-års perspektiv?	**Interviewer**: how do you expect the acquisition to affect Careto's key figures with a 2-year perspective?
Velkalibrert Tor/Berit: Jeg håper at vi når en markedsandel på privatmarkedet på 25 %, opp fra dagens nivå på 20 %. Altså en økning med 5% fra dagens andel. I tillegg håper jeg at utlånsinntektene fra privatmarkedet vil øke fra dagens nivå på 10 milliarder til 12 milliarder. Altså en utlånsvekst på 20 %. Jeg mener det er realistisk å tro at vi skal få til dette.	**Well-calibrated Tor/Berit:** I hope that we will reach a market share on the private market of 25 %, up from the current level of 20 %, representing an increase of 5 % from the current share. In addition, I hope that lending income from the private market will increase from the current level of 10 billion to 12 billion. In other words, a lending growth of 20 %. I think it is realistic to believe that we will achieve this

Appendix F

Images of the Female and Male CEO

Note. Images were purchased from the online stock image provider, Shutterstock.

Original version (Norwegian)

Note. Fictitious annual reports excerpts for the well-calibrated condition

Note. Fictitious annual reports excerpts for the overconfident condition

1 Lending revenues PM (in bn. NOK)

The objective of the acquisition of Ybank was to increase lending revenues in the private market (PM) by 20%. Careto has achieved a 21% increase, **1%** <u>more</u> than estimated.

■ Achieved lending revenues
▨ Estimated lending revenues

+21 %

2018   2019   2020   2021 (10,0)   2022 (12,1 12,0)

2 Market share PM (in percent)

The objective of the acquisition of Ybank was an increased market share in the private market (PM) by 5 percentage points. Careto has achieved a 5 percentage point increase, exactly <u>as</u> estimated.

25% →+5   25% ⤍ +5
20   20

■ Achieved market share   ▨ Expected market share
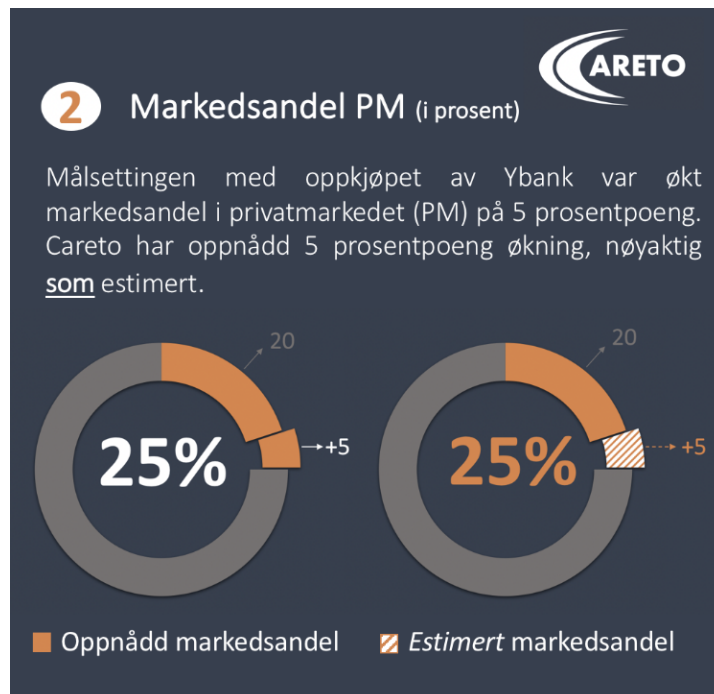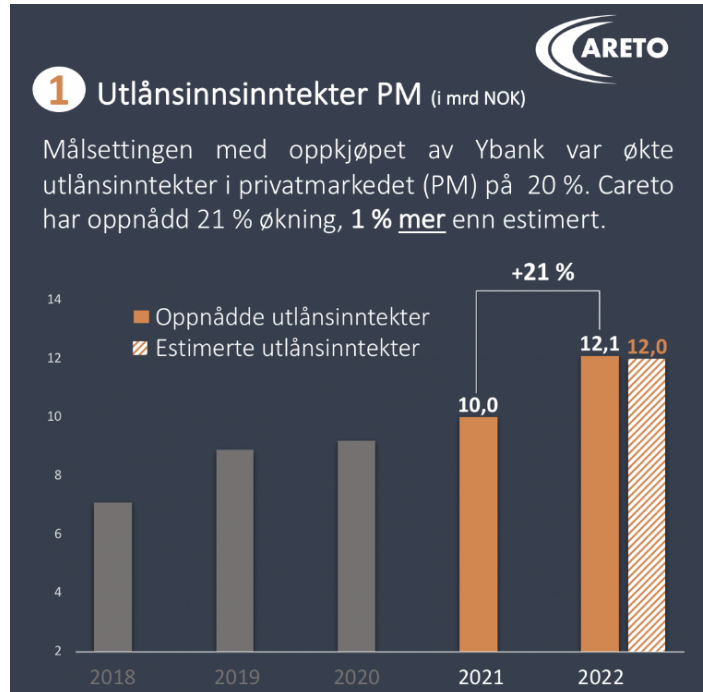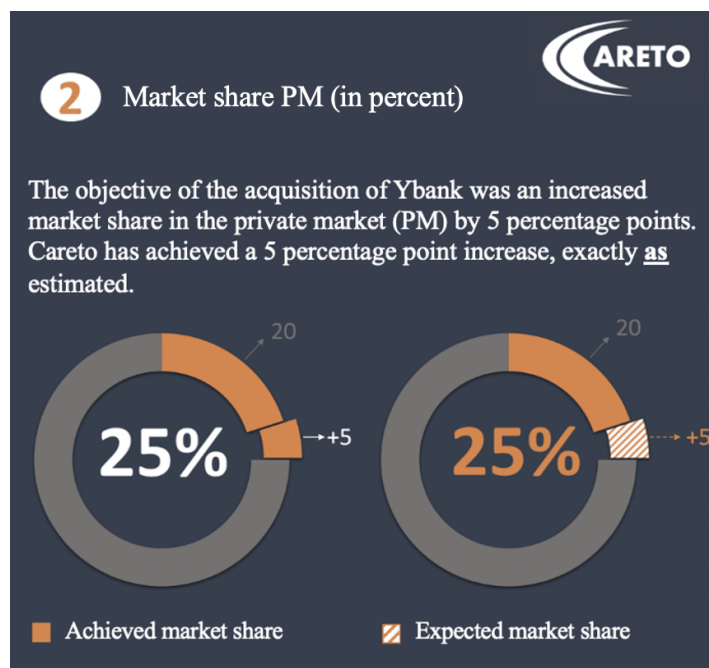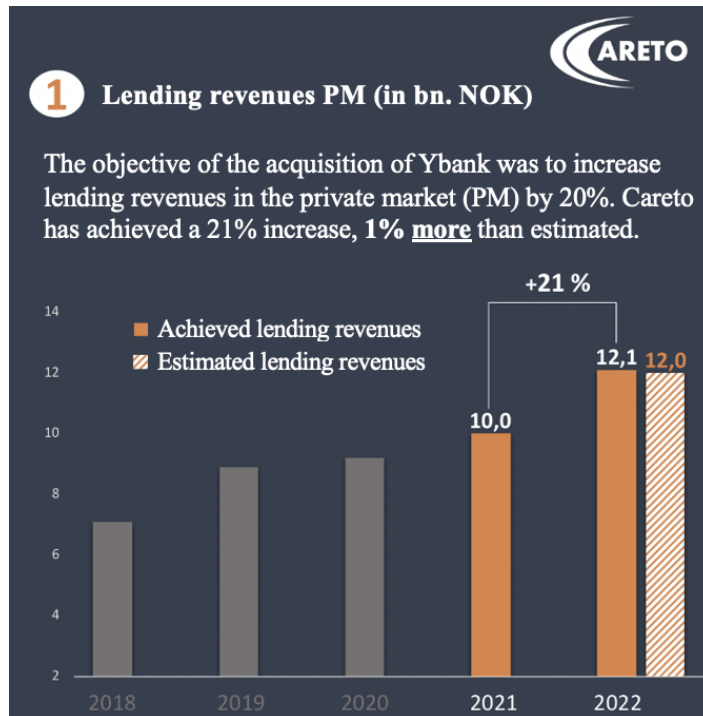
Note. Fictitious annual reports excerpts for the well-calibrated condition
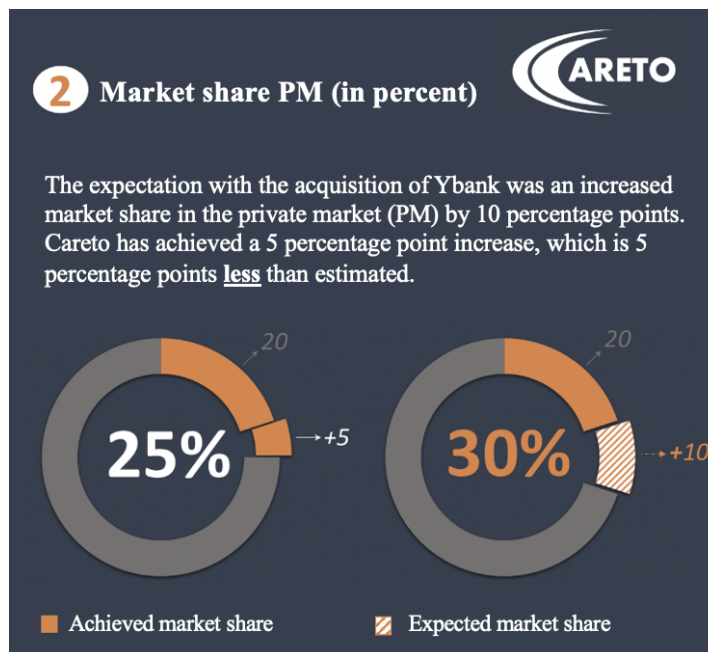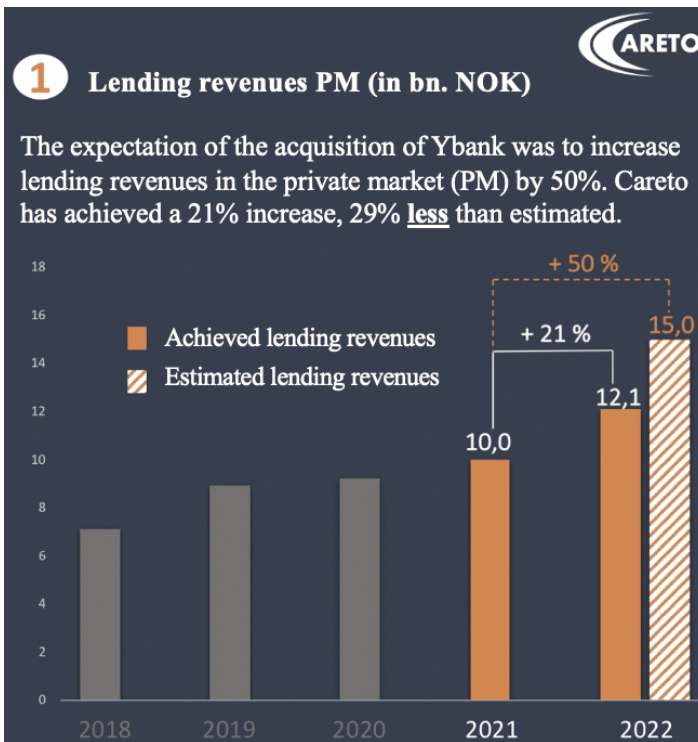
Note: Fictitious annual reports excerpts for the overconfident condition

Appendix H

Histograms for Competence and Warmth for Phase 1 and Phase 2

Competence Phase 1

Note: Mean = 3.68, Std.Dev = .775, *N* = 454

Competence Phase 2

Note: Mean = 3.58, Std.Dev = .873, *N* = 454

Warmth Phase 1

Note: Mean = 2.81, Std.Dev = .863, *N* = 454

Warmth Phase 2

Note: Mean = 2.82, Std.Dev = .89, *N* = 454
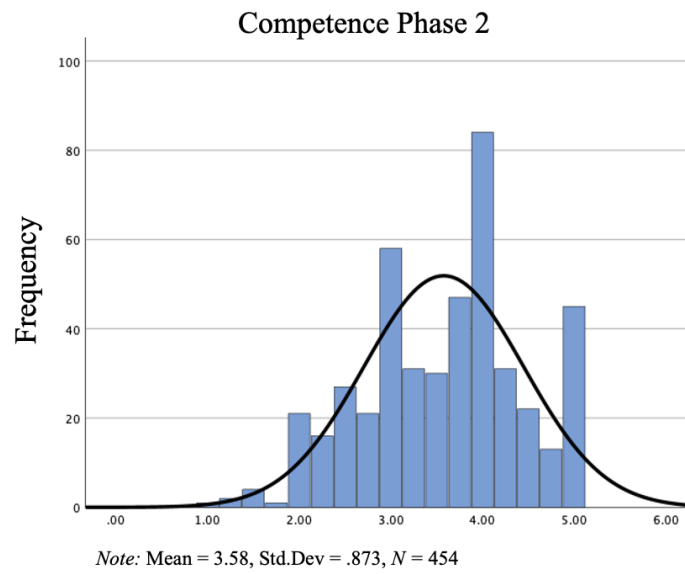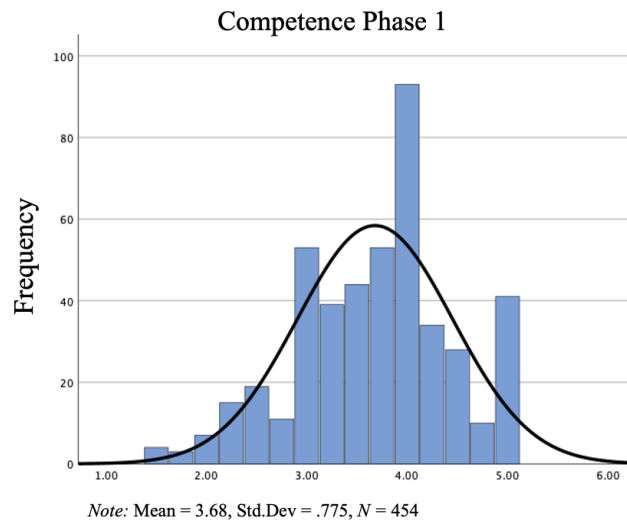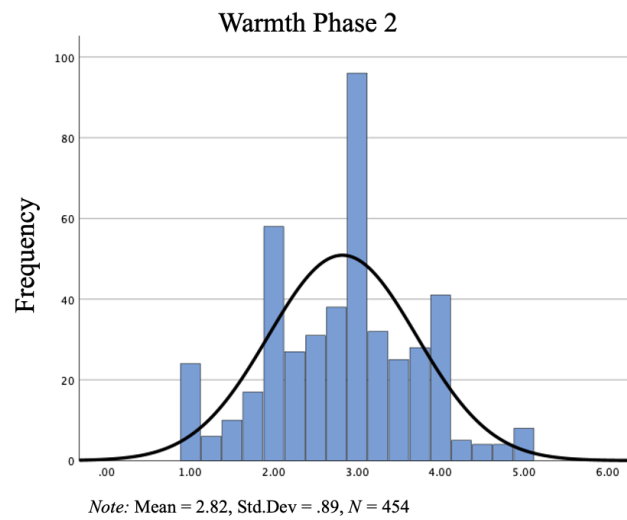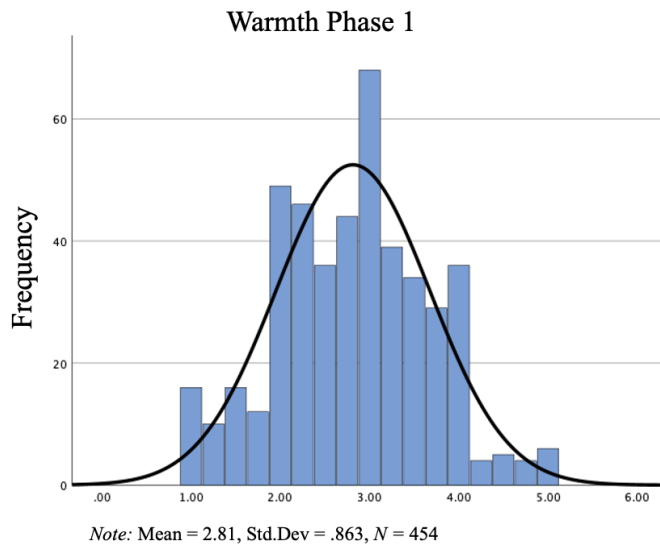
Skewness and Kurtosis for Competence and Warmth

	Mean Statistic	St. Deviation	Skewness		Kurtosis	
			Statistics	Std. Error	Statistics	Std. Error
Competence Phase 1	3.6790	.77528	-.323	.115	-.155	.229
Competence Phase 2	3.5771	.87291	-.260	.115	.-497	.229
Warmth Phase 1	2.8128	.86262	.024	.115	-.290	.229
Warmth Phase 2	2.8232	.88876	-.022	.115	-.244	.229

Note. N=454

Appendix I

Q-Q Plots for Competence and Warmth in Phase 1 and Phase 2

Normal Q-Q Plots of Competence Phase 1

Normal Q-Q Plots of Warmth Phase 1

Normal Q-Q Plots of Competence Phase 2

Normal Q-Q Plots of Warmth Phase 2
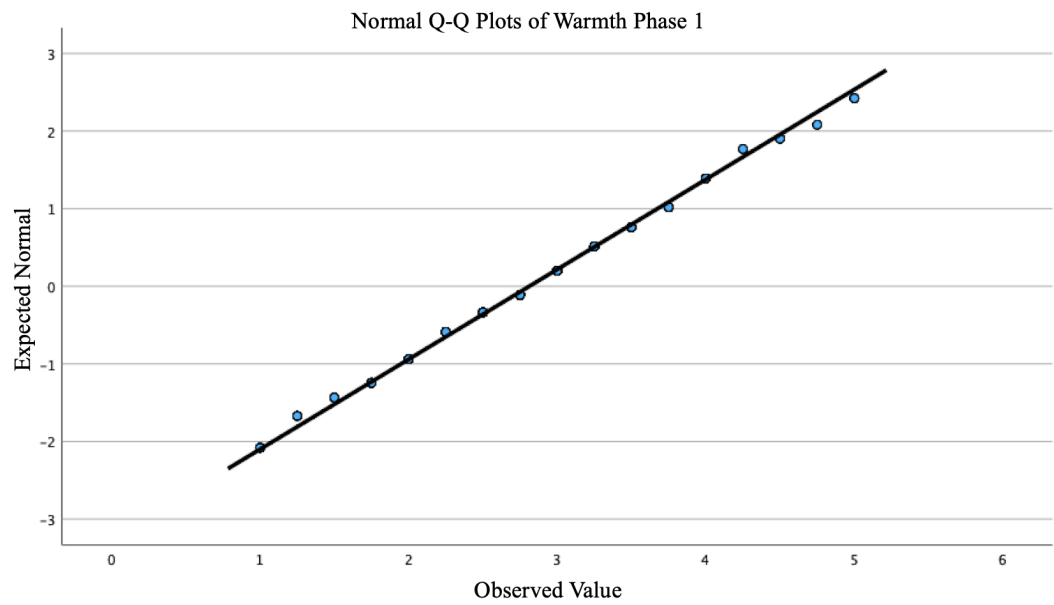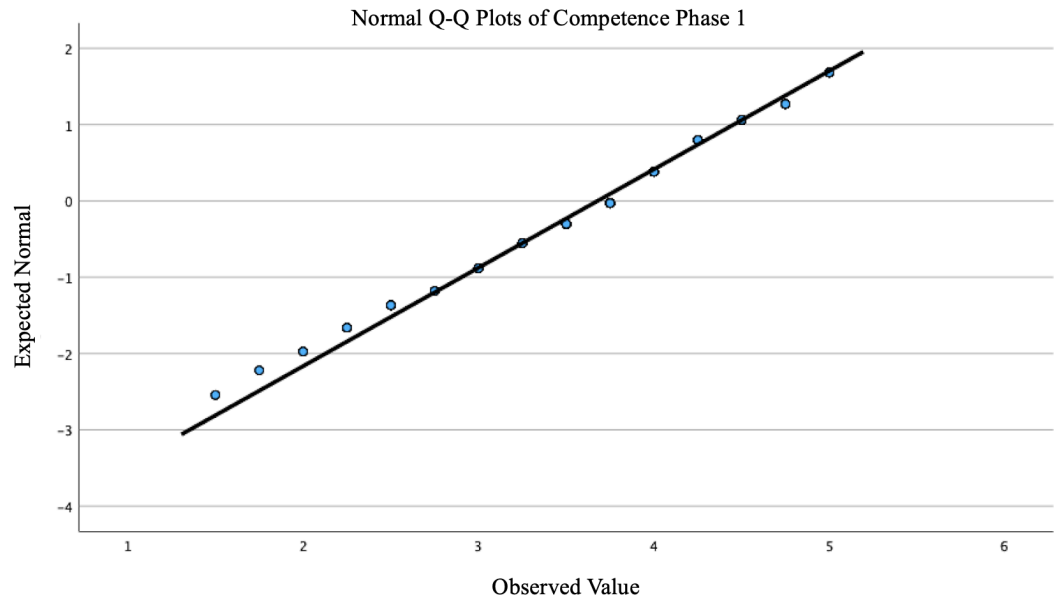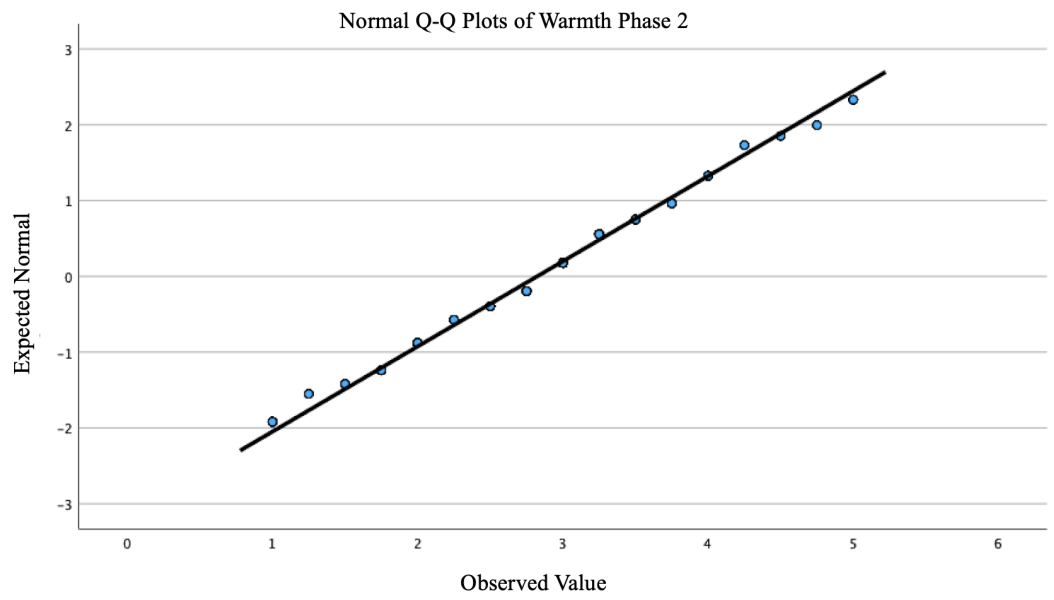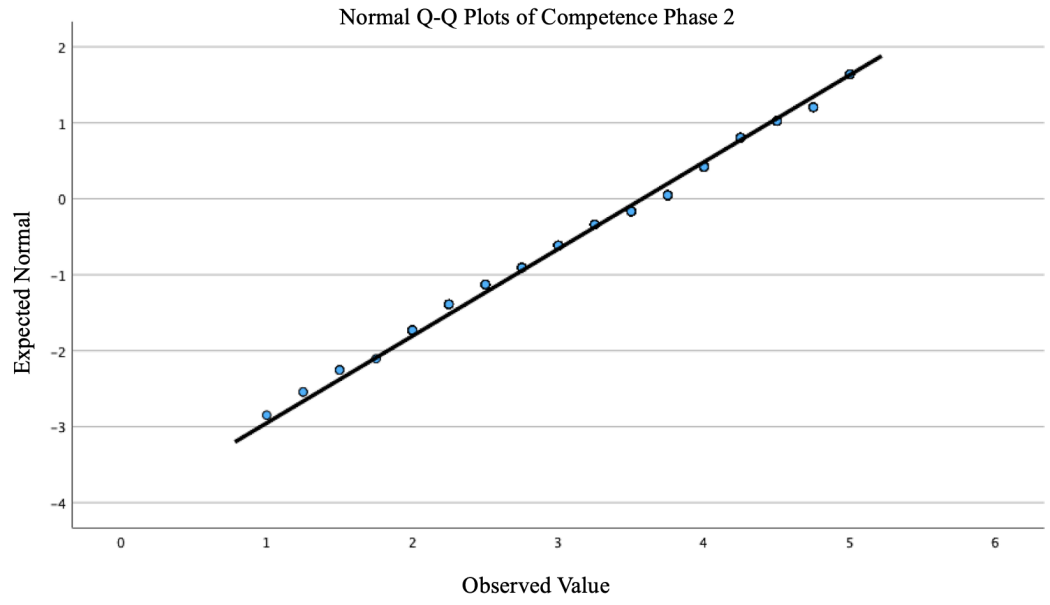
Appendix J

Levene's Test of Equality of Error Variances

	Levene Statistic	df1	df2	Sig.
Competence Phase 1	1.232	3	450	.298
Competence Phase 2	3.324	3	450	.020
Warmth Phase 1	.179	3	450	.911
Warmth Phase 2	.190	3	450	.903

Note. N=454

Appendix K

Competence Ratings Phase 1 and Phase 2

Descriptive Statistics

Dependent Variable: Competence Phase 1

Source	Confidence Level	Mean	St. Deviation	N
Male	Calibrated	3.4818	.77052	110
	Overconfident	3.3606	.75588	113
	Total	3.4202	.76384	223
Female	Calibrated	3.9457	.66291	115
	Overconfident	3.9116	.74181	116
	Total	3.9286	.70231	231
Total	Calibrated	3.7189	.75270	225
	Overconfident	3.6397	.79652	229
	Total	3.6790	.77527	454

Test of Between-Subject Effects

Dependent Variable: Competence Phase 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	30.186a	3	10.062	18.704	<.001	.111
Intercept	6128.802	1	6128.802	11392.507	<.001	.962
CEOGender	29.212	1	29.212	54.301	<.001	.108
Confidence Level	.683	1	.683	1.270	.260	.003
CEOGender * Confidence Level	.216	1	.216	.401	.527	.001
Error	242.086	450	.583			

Total	6417.063	454
Corrected Total	272.272	453

Note. R squared = .111 (Adjusted R squared = .105)

Descriptive Statistics

Dependent Variable: Competence Phase 2

Source	Confidence Level	Mean	St. Deviation	N
Male	Calibrated	3.8545	.70498	110
	Overconfident	2.8363	.69498	113
	Total	3.3386	.86489	223
	Calibrated	4.1674	.65621	115
	Overconfident	3.4504	.80974	116
	Total	3.8074	.81874	231
Total	Calibrated	4.0144	.69682	225
	Overconfident	3.1474	.81404	229
	Total	3.5771	.87291	454

Test of Between-Subject Effects

Dependent Variable: Competence Phase 2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	112.415a	3	37.472	72.444	<.001	.326
Intercept	5807.034	1	5807.034	11226.786	<.001	.961
CEOGender	24.373	1	24.373	47.121	<.001	.095
Confidence Level	85.402	1	85.402	165.108	<.001	.268
CEOGender *	2.575	1	2.575	4.978	.026	.011

Confidence Level			
Error	232.762	450	.517
Total	6154.375	454	
Corrected Total	345.177	453	

Note. R squared = . 326 (Adjusted R squared = .321)

Pairwise Comparisons (Bonferroni)

Dependent variable: Competence Phase 2

						95% Confidence Interval for difference	
CEOGender	Confidence Level	Confidence Level	Mean Difference (I-J)	Std. Error	Sig.b	Lower Bound	Upper Bound
Male	Calibrated	Overconfident	1.018*	.096	<.001	.829	1.208
	Overconfident	Calibrated	-1.018*	.096	<.001	-1.208	-.829
Female	Calibrated	Female	.717*	.096	<.001	.531	.903
	Overconfident	Female	-.717	.096	<.001	-.903	-.531

Note. Based on estimated marginal means

*. The mean difference is significant at the -05 level

b. Adjustment for multiple comparisons: Bonferroni

Appendix L

Competence Ratings Across Time Of Measurement

Test of Within-Subject

Dependent Variable: Competence

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time	2.169	1	2.169	14.445	<.001	.031
Time*CEOGender	.109	1	.109	.729	.394	.002
Time*ConfidenceLevel	35.404	1	35.404	235.788	<.001	.344
Time*CEOGender * ConfidencLevel	.650	1	.640	4.330	.038	.010
Error	67.567	450	.150			

Note. Time = Time of Measurement

Test of Within-Subject

Dependent Variable: Warmth

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time	.026	1	.026	.554	.457	.001
Time*CEOGender	.002	1	.002	.039	.843	.000
Time*ConfidenceLevel	.321	1	.321	6.817	.009	.015
Time*CEOGender * ConfidencLevel	.184	1	.184	3.902	.049	.009
Error	21.179	450	.047			

Note. Time = Time of Measurement

Pairwise Comparisons (Bonferroni)

Competence

CEOGender	Confidence Level	Time (I)	Time (J)	Mean Difference (I-J)	Std. Error	Sig.b	95% Confidence Interval for difference	
							Lower Bound	Upper Bound
Male	Calibrated	1	2	-.373*	.052	<.001	-.475	-.270
		2	1	.373*	.052	<.001	.270	.475
	Overconfident	1	2	.524*	.052	<.001	.423	.626
		2	1	-.524*	.052	<.001	-.626	-.423
Female	Calibrated	1	2	-.222*	.051	<.001	-.322	-.121
		2	1	.222*	.051	<.001	.121	.322
	Overconfident	1	2	.461*	.051	<.001	.361	.561
		2	1	-.461*	.051	<.001	-.561	-.361

Note. Based on estimated marginal means

*. The mean difference is significant at the -05 level

b. Adjustment for multiple comparisons: Bonferroni

Pairwise Comparisons (Bonferroni)

Warmth

CEOGender	Confidence Level	Time (I)	Time (J)	Mean Difference (I-J)	Std. Error	Sig.b	95% Confidence Interval for difference	
							Lower Bound	Upper Bound
Male	Calibrated	1	2	-.023	.029	.438	-.080	.035
		2	1	.023	.029	.438	-.035	.080
	Overconfident	1	2	-.004	.029	.878	-.061	.052
		2	1	.004	.029	.878	-.052	.061
Female	Calibrated	1	2	-.074*	.029	.010	-.130	-.018
		2	1	.074*	.029	.010	.018	.130

Overconfident	1	2	.058*	.029	.042	.002	.114
	2	1	-.058	.029	.042	-.114	-.002

Note. Based on estimated marginal means

    *. The mean difference is significant at the -05 level

    b. Adjustment for multiple comparisons: Bonferroni

Appendix M

Additional Analyses – Warmth Ratings

Descriptive Statistics

Dependent Variable: Warmth Phase 1

Source	Confidence Level	Mean	St. Deviation	N
Male	Calibrated	2.7545	.85267	110
	Overconfident	2.4358	.78617	113
	Total	2.5930	.83324	223
	Calibrated	3.1065	.81250	115
	Overconfident	2.9440	.85978	116
	Total	3.0249	.83872	231
Total	Calibrated	2.9344	.84903	225
	Overconfident	2.6932	.86098	229
	Total	2.8128	.86262	454

Test of Between-Subject Effects

Dependent Variable: Warmth Phase 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	28.347a	3	9.449	13.773	<.001	.084
Intercept	3583.910	1	3583.910	5223.708	<.001	.921
CEOGender	20.982	1	20.982	30.583	<.001	.064
ConfidenceLevel	6.569	1	6.569	9.575	.002	.021
CEOGender * ConfidenceLevel	.692	1	.692	1.008	.316	.002
Error	308.738	450	.686			
Total	3929.000	454				
Corrected Total	337.086	453				

Note. R squared = .084 (Adjusted R squared = .078)

Appendix N

Additional Analyses – Participant Gender On Competence Ratings at Phase 1 and Phase 2

Test of Between-Subjects Effects

Dependent Variable: Competence Phase 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	43.849a	7	6.264	12.231	<.001	.161
Intercept	6096.674	1	6096.674	11903.885	<.001	.964
CEOGender	29.695	1	29.695	57.981	<.001	.115
ConfidenceLevel	.502	1	.502	.979	.323	.002
RaterGender	9.498	1	9.498	18.545	<.001	.040
CEOGender * ConfidenceLevel	.136	1	1.850	.266	.606	.001
CEOGender * RaterGender	1.850	1	1.898	3.611	.058	.008
ConfidenceLevel* RaterGender	1.898	1	.161	3.705	.055	.008
CEOGender*ConfidenceLevel*RaterGender	.161	1	.512	.315	.575	.001
Error	228.423	446				
Total	6418.063	454				
Corrected Total	272.272	453				

Note. R squared = .161 (Adjusted R squared = .148)

Pairwise Comparisons (Bonferroni)

Dependent variable: Competence

CEOGender	ConfidenceLevel	(I) Rater Gender	(J) Rater Gender	Mean Difference (I-J)	Std. Error	Sig.b	95% Confidence Interval for difference	
							Lower Bound	Upper Bound
Male	Calibrated	Male	Female	.005	.138	.969	-.266	.276
		Female	Male	-.005	.138	.969	-.276	.266
	Overconfident	Male	Female	-.331*	.136	.015	-.597	-.064
		Female	Male	.331*	.136	.015	.064	.597
Female	Calibrated	Male	Female	-.327*	.134	.015	-.591	-.064
		Female	Male	.327*	.134	.015	.064	.591
	Overconfident	Male	Female	-.512*	.133	<.001	-.774	-.250
		Female	Male	.512*	.134	<.001	.250	.774

Note. Based on estimated marginal means

    *. The mean difference is significant at the -05 level

    b. Adjustment for multiple comparisons: Bonferroni

Test of Between-Subject Effects

Dependent Variable: Competence Phase 2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	118.574a	7	16.939	33.340	<.001	.344
Intercept	5758.712	1	5758.712	11334.292	<.001	.962
CEOGender	24.903	1	24.903	49.014	<.001	.099
ConfidenceLevel	83.448	1	83.448	164.243	<.001	.269
RaterGender	3.256	1	3.256	6.409	0.12	.014
CEOGender * ConfidenceLevel	2.644	1	2.655	5.204	.023	.012
CEOGender * RaterGender	1.619	1	1.619	3.187	.075	.007

	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
ConfidenceLevel*Rater Gender	.709	1	.709	1.395	.238	.003
CEOGender*Confidence Level*RaterGender	.408	1	.408	.803	.371	.002
Error	226.603	446	.508			
Total	6154.375	454				
Corrected Total	345.177	453				

Note. R squared = .344 (Adjusted R squared = .333)

Pairwise Comparisons (Bonferroni)

Dependent variable: Competence

CEOGender	ConfidenceLevel	(I) Rater Gender	(J) Rater Gender	Mean Difference (I-J)	Std.Error	Sig.b	95% Confidence Interval for difference	
							Lower Bound	Upper Bound
Male CEO	Calibrated	Male	Female	-.031	.137	.821	-.301	.239
		Female	Male	.031	.137	.821	.-239	.310
	Overconfident	Male	Female	-.069	.135	.607	-.335	.196
		Female	Male	.069	.135	.607	-.196	.335
Female CEO	Calibrated	Male	Female	-.151	.134	.259	-.413	.112
		Female	Male	.151	.134	.259	-.112	.413
	Overconfident	Male	Female	-.431	.133	.001	-.691	-.17
		Female	Male	.431	.133	.001	.170	.691

Note. Based on estimated marginal means

*. The mean difference is significant at the -05 level

b. Adjustment for multiple comparisons: Bonferroni

Appendix O

Additional Analyses – Leader Optimism and Certainty

Correlation – Calibrated Condition

		Certainty	Optimism	Competence Phase 1	Warmth Phase 1	Competence Phase 2	Warmth Phase 2
Certainty	Pearson Correlation	1	.264**	1.184**	.059	.152*	.039
	Sig. (2-tailed)		<.001	.006	.378	.022	.564
	N	225	225	225	225	225	225
Optimism	Pearson Correlation	.264**	1	.067	.080	.086	.033
	Sig. (2-tailed)	<.001		.317	.232	.200	.625
	N	225	225	225	225	225	225
Competence Phase 1	Pearson Correlation	.184**	.067	1	.434**	.797**	.439**
	Sig. (2-tailed)	.006	.317		<.001	<.001	<.001
	N	225	225	225	225	225	225
Warmth Phase 1	Pearson Correlation	.059	.080	.434**	1	.421**	.929**
	Sig. (2-tailed)	.378	.232	<.001		<.001	<.001
	N	225	225	225	225	225	225
Competence Phase 2	Pearson Correlation	.152*	.086	.797**	.421**	1	.482**
	Sig. (2-tailed)	.022	.200	<.001	<.001		<.001
	N	225	225	225	225	225	225
Warmth Phase 2	Pearson Correlation	.039	.033	.439**	.092**	.482**	1
	Sig. (2-tailed)	.564	.625	<.001	<.001	<.001	
	N	225	225	225	225	225	225

Note. ** Correlation is significant at the .01 level (2-tailed). *Correlation is significant at the .05 level (2-tailed). a. ConfidenceLevel = Calibrated

Correlation – Overconfident Condition

		Certainty	Optimism	Competence Phase 1	Warmth Phase 1	Competence Phase 2	Warmth Phase 2
Certainty	Pearson Correlation	1	.453**	.232**	.288**	.292**	.274**
	Sig. (2-tailed)		<.001	<.001	<.001	<.001	<.001
	N	229	229	229	229	229	229
Optimism	Pearson Correlation	.453**	1	.248**	.209**	.282**	.171**
	Sig. (2-tailed)	<.001		<.001	.001	<.001	.010
	N	229	229	229	229	229	229
Competence Phase 1	Pearson Correlation	.232**	.248**	1	.576**	.702**	.509**
	Sig. (2-tailed)	<.001	<.001		<.001	<.001	<.001
	N	229	229	229	229	229	229
Warmth Phase 1	Pearson Correlation	.288**	.209**	.576**	1	.587**	.945**
	Sig. (2-tailed)	<.001	.001	<.001		<.001	<.001
	N	229	229	229	229	229	229
Competence Phase 2	Pearson Correlation	.292**	.282**	.702**	.587**	1	.597**
	Sig. (2-tailed)	<.001	<.001	<.001	<.001		<.001
	N	229	229	229	229	229	229
Warmth Phase 2	Pearson Correlation	.274**	.171**	.509**	.945**	.597**	1
	Sig. (2-tailed)	<.001	.010	<.001	<.001	<.001	
	N	229	229	229	229	229	229

Note. ** Correlation is significant at the .01 level (2-tailed). *Correlation is significant at the .05 level (2-tailed). a. ConfidenceLevel = Calibrated