GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Emergence of Leadership:

Predicting Leadership Initiative and its Success from Individual Traits, Motives, and Characteristics

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades „*Doctor rerum naturalium*“

der Georg-August-Universität Göttingen

im Promotionsprogramm Behavior and Cognition

der Georg-August University School of Science (GAUSS)

vorgelegt von

Johanna Prüfer

aus Paderborn

Göttingen, 2019

Betreuungsausschuss

Prof. Dr. Stefan Schulz-Hardt, Abteilung Wirtschafts- & Sozialpsychologie, Georg-Elias-Müller-Institut für Psychologie, Georg-August-Universität Göttingen

Prof. Dr. Andreas Mojzisch, Institut für Psychologie, Universität Hildesheim

Dr. Dietmar Zinner, Kognitive Ethologie, Deutsches Primatenzentrum

Mitglieder der Prüfungskommission

Referent: Prof. Dr. Stefan Schulz-Hardt, Abteilung Wirtschafts- & Sozialpsychologie, Georg-Elias-Müller-Institut für Psychologie, Georg-August-Universität Göttingen

Korreferent: Prof. Dr. Andreas Mojzisch, Institut für Psychologie, Universität Hildesheim

Weitere Mitglieder der Prüfungskommission

Dr. Dietmar Zinner, Kognitive Ethologie, Deutsches Primatenzentrum

Prof. Dr. Lars Penke, Abteilung Biologische Persönlichkeitspsychologie, Georg-Elias-Müller- Institut für Psychologie, Georg-August-Universität Göttingen

Prof. Dr. Margarete Boos, Abteilung Sozial- und Kommunikationspsychologie, Georg-Elias-Müller-Institut für Psychologie, Georg-August-Universität Göttingen

Prof. Dr. Annekathrin Schacht, Abteilung Affektive Neurowissenschaft und Psychophysiologie, Georg-Elias-Müller- Institut für Psychologie, Georg-August-Universität Göttingen

Tag der mündlichen Prüfung: 22.05.2019

Danksagung

Während meiner Promotion haben mich viele Menschen begleitet und ich kann mich glücklich schätzen, so vielen Menschen danken zu können.

Zunächst gilt mein aufrichtiger Dank meinem Doktorvater Stefan Schulz-Hardt, der mir nicht nur fachlich mit seiner Expertise zur Seite stand, sondern mir die Möglichkeit gegeben hat, in großem Maß selbstständig an meinem Promotionsprojekt zu arbeiten und mich persönlich maßgeblich weiterzuentwickeln.

Ich danke meinen Thesis Committee Mitgliedern Andreas Mojzisch und Dietmar Zinner für das kontinuierliche Interesse und Unterstützung in meinem Forschungsprojekt.

Margarete Boos, Annekathrin Schacht und Lars Penke für die Bereitschaft, Teil meines Promotionskomittees zu sein.

Dank gebührt auch meinen Kollegen, die mir stets mit fachlichem Rat und Unterstützung zur Seite gestanden haben.

Meinen Hilfskräften, Nicolina, Birk, Julian, Jaro, Kira und Yannick, meinen Praktikanten Katharina und Nadya, allen Hilfskräften, die in meinem Projekt ausgeholfen haben und allen Bachelor- und Masterabsolventen, ohne die dieses Forschungsprojekt nicht möglich gewesen wäre. Durch euren Einsatz und der hohen Qualität eurer Arbeit habt ihr maßgeblich zum Gelingen dieses Projekts beigetragen und mir außerdem in der Zusammenarbeit viel Freude bereitet.

Ich möchte mich an dieser Stelle auch herzlich bei Paul bedanken, der mir mit statistischem Rat durch die Promotion hinweg ausgeholfen hat.

Besonders möchte ich auch meinen Freunden hier in Göttingen und in Paderborn danken, Julia, Jana, Moritz, Fede, Sarah, Lena, Chrissi, Flo, Johanna, Jana und Lisa, die während meiner Promotion zwar oft zurückstecken mussten, die aber immer für mich da sind.

Besonders möchte ich mich an dieser Stelle bei Isi bedanken, die mich stets bestärkt und motiviert hat und an mich geglaubt hat, wenn ich es nicht getan habe.

Großer Dank gilt meiner großen Familie, die mir immer zur Seite steht und ohne die ich gar nichts wäre.

Zuletzt danke ich Christian, ohne den diese Promotion niemals zustande gekommen und möglich gewesen wäre. Deine Unterstützung bedeutet mir alles.

Göttingen, April 2019 *Johanna Prüfer*

Table of Contents

Summary

Research within the trait-oriented approach to leadership has focused more on leader effectiveness and less so on the emergence of leadership. Furthermore, most studies in the latter tradition have relied on self-descriptions or follower ratings of leadership emergence, which allowed for possible confounds with implicit leadership theories. In the present research, we propose a new approach to emergence of leadership by differentiating between leadership initiative, on the one hand, and its success in attracting followers, on the other hand. We test this approach by using strictly behavioral measures of both components in a laboratory setting. In two large studies, we aim at identifying inter-individual differences to predict these two components with predictors already known from the literature, like extraversion, as well as new predictors, like testosterone. In Study 1, $N = 392$ students took part in two laboratory sessions. During the first session, we collected data on most of the predictor variables. In the second session, participants worked on different estimation tasks (e.g., estimating a person's body height) in ad-hoc groups of four. In Study 2, $N = 380$ students completed the first session online. In the second session, they worked in ad-hoc groups of four on wilderness survival tasks (e.g., deciding on how to cross a river) in the laboratory. In both studies, participants either worked in mixed-gender, male-only, or female-only groups. Analyses aggregating data from both studies suggest that extraversion and participants' self-reported subjective competence robustly predict leadership initiative but not its success, while general intelligence predicts both components of leadership emergence. While the effects of differences in extraversion and general intelligence on leadership initiative are consistent with previous research, we were unable to replicate previously reported effects of other inter-individual differences (e.g., narcissism) on emergence of leadership. Based on our findings, we discuss implications and future research questions.

Introduction

What exactly is leadership? Within leadership research, it is well known that the search for one universal definition of leadership is never ending (Bass & Bass, 2008) and there seem to be as many definitions of leadership as there are attempts to define it (Stogdill, 1974). However, most definitions share a common core, namely "influencing others" (Yukl, 2013). Although it is applied to organizations, the Global Leadership, and Organizational Behavior Effectiveness Research project (GLOBE) provides a comprehensive definition of leadership. According to GLOBE, leadership is "the ability of an individual to influence, motivate, and enable others to contribute towards the effectiveness and success of the organizations of which they are members" (House, Hanges, Javidan, Dorfman, & Gupta, 2004, p. 15).

In this context, how leadership emerges poses an interesting question. Research on leadership emergence aims at investigating the factors that determine who will emerge as a leader in a group (Crockett, 1955). The present research has the goal to deepen the understanding of emergence of leadership. Previous research considered emergence of leadership as a global phenomenon while neglecting whether the emerging leader is successful in his leadership attempt. Therefore, we differentiate emergence of leadership into two components, namely leadership initiative and its success in attracting followers. Because it remains unclear whether previously identified predictors for leadership emergence predict these two newly introduced components differentially, we use a broad range of inter-individual traits, motives, and characteristics for their prediction. Moreover, previous research mainly used subjective measures of leadership emergence, which might reflect raters' implicit leadership theories rather than who is actually showing leadership behavior. Therefore, we aim at measuring these two components behaviorally. In the following, we will first provide a brief historical overview of research on leadership and how leadership is

measured before giving an overview of the research and already identified predictors for emergence of leadership. Subsequently, we describe two studies we conducted. Finally, we discuss our results and derive implications for future research.

Research on Leadership – a Brief Historical Overview

Great-Man and Trait Theory of Leadership. Leadership seems to be a phenomenon that caught people's and researchers' interest very early at the beginning of psychological research. Research approaches on leadership began with the idea that history is shaped by the leadership of great men (Bass & Bass, 2008), who are superior individuals and essentially born as leaders (Kirkpatrick & Locke, 1991). Back then, Napoleon was a prominent example of a "great man" influencing history. Today, you might consider a successful corporate leader as a great man, like Tesla's CEO Elon Musk or the late Steve Jobs (Bass & Bass, 2008). Leadership research in the 19th and the early 20th century was dominated by the "Great-Man Theory" (Kirkpatrick & Locke, 1991). This approach concentrates on the leader as an individual, and therefore paved the way for the Trait Theory of Leadership, focusing on specific qualities of a leader. Its main idea is that leaders are different from followers, possessing different characteristics (Bass & Bass, 2008; Colbert, Judge, Choi, & Wang, 2012), for example, being very masculine. Historically, the trait approach to leadership seems to fall in and out of favor. Its popularity seemed to end when influential reviews concluded that there were no reliable correlations between a person's characteristics and leadership (Mann, 1959; Stogdill, 1948). Despite some skepticism among leadership researchers, the trait approach to leadership regained interest in terms of literature reviews, re-analyses of existing data, meta-analyses, and new research, for example on more contemporary traits like narcissism (e.g., Brunell et al., 2008; Judge, Bono, Ilies, & Gerhardt, 2002; Lord, De Vader, & Alliger, 1986).

Behavioral approaches. Shortly after the interest in leader characteristics temporarily diminished, researchers took a closer look at how leaders *behave*. Most influential in this research avenue were the Ohio State Leadership Studies, identifying two fundamental leader behaviors, that is, consideration (the extent to which a leader expresses concern for his group members, by, for example, including followers in decisions) and initiating structure (the extent to which a leader initiates, organizes, and structures work in the group, for example, insisting on meeting deadlines; Bass & Bass, 2008; Fleishman, 1953). Although leadership research was subsequently dominated by consideration and initiating structure, the research on these constructs contained several weaknesses, for example, a lack of identifying causal relationships between these and methodological shortcomings, for example, common method bias, leading researchers in the 1970s to investigate situational effects, which were largely neglected before (Judge, Piccolo, & Ilies, 2004b).

Contingency theory. Contingency theories introduced the idea of situational effects to leadership research and focused on how these effects moderate the influence of leaders' traits or behaviors on leader effectiveness (Bass & Bass, 2008; Yukl, 2013). The path-goal theory of leader effectiveness is one of the most prominent contingency theories (House, 1971, 1996). The theory's main idea is that leaders should adapt their leadership styles to their subordinates' work environment and abilities by clarifying the way to attain the goal, ensuring that the subordinates expect to reach it, experience intrinsic motivation, and receive positive valent rewards when they reach the goal. In this way, the theory addresses how leaders affect their subordinates' motivation and satisfaction (House, 1996). Within this approach, several classes of leader behaviors are outlined, including path-goal clarifying behaviors, achievement-oriented behaviors, supportive behaviors, and participative behaviors. However, like most contingency theories of leadership, the theory received only mixed empirical support (House, 1996).

Transactional, transformational, and charismatic leadership theory. Transactional, transformational, and charismatic leadership theories focus on which benefits leaders and followers can provide for each other (Conger & Kanungo, 1998). As the most basic form, transactional leadership can be understood as a mere exchange of resources. In contrast, transformational leaders hold out transcendent aims and therefore address the followers' higher-order needs (Conger & Kanungo, 1998; Judge & Piccolo, 2004). As described by Judge and Piccolo (2004), transformational leadership consists of four dimensions, namely charisma or idealized influence, inspirational motivation, intellectual stimulation, and individualized consideration. As charisma is already included as one on the four dimensions, it becomes evident here that, transformational and charismatic leadership overlap, and they even complement each other (Judge & Piccolo, 2004). Although they represent different approaches, they are often used interchangeably (Yukl, 2013). A comprehensive meta-analysis (Judge & Piccolo, 2004) on transformational and transactional leadership confirmed the overall validity and quality of transformational leadership but also showed that specific forms of transactional leadership performed as well as transformational leadership.

The presented approaches to leadership research are concerned with factors influencing leadership. In this context, it is important what exactly is influenced, and therefore, how leadership is operationalized. In the following, we will present how leadership is typically measured.

Leadership Criteria

In general, leadership can be measured twofold. On the one hand, we can measure what makes leadership successful, that is, a leader's effectiveness. On the other hand, we can examine how someone evolves as a leader in the perception of others, that is, leadership

emergence (Lord et al., 1986). In the following, we will take a closer look at these two leadership criteria.

Leadership effectiveness. Effective leadership is defined as a leader's success in influencing his followers for them to succeed in reaching their goals (Bass & Bass, 2008). Hence, leadership effectiveness refers to the leader's influence on an organization's success, for example, the profitability of a department (Hogan, Curphy, & Hogan, 1994). Ideally, leadership effectiveness is measured as either team, group, or organizational effectiveness (Hogan et al., 1994), whereas, in most cases, actual assessments of leadership effectiveness come from the leaders' supervisors, peers, subordinates, or a combination of them (Judge et al., 2002). In meta-analyses investigating leadership effectiveness, the authors typically use measures of follower motivation, follower satisfaction, and performance among the subordinates to assess leadership effectiveness (Judge et al., 2002, 2004b). Leadership effectiveness represents a between-groups phenomenon concerning the ability of a leader to influence her subordinates. It is important to note that in order to evaluate a leader's effectiveness, this person must already have a leadership position (Judge et al., 2002). Besides evaluating a leader's success in influencing others to reach goals, leadership can be investigated as to how someone evolves as a leader, that is, emergence of leadership.

Emergence of leadership. Judge and colleagues (2002) point to the fact that leadership emergence is a phenomenon only occurring within groups, as leadership can only emerge among one person becoming a leader and at the same time other persons who are led by this leader as followers. The same authors also state that "leader emergence refers to whether (or to what degree) an individual is viewed as a leader by others, who typically have only limited information about that individual's performance" (p. 767). The present research takes this definition as a basis for leadership emergence. We want to point out here, as can be seen from the quote, that emergence of leadership is traditionally understood as the

perceptions others have of the emerging leader, that is, whether these others perceive the person as “leader-like” (Hogan et al., 1994).

Compared to each other, research within the trait-oriented approach to leadership has mostly focused on leadership effectiveness and less so on leadership emergence (e.g., Judge et al., 2002). In order to compensate for this disadvantage, we focus on research on emergence of leadership.

Research on Emergence of Leadership

In the following, we want to provide a brief outline on how research on emergence of leadership is conducted, before giving an overview of previous findings from that research field and subsequently stating our research aims.

Research on emergence of leadership aims at identifying the factors deciding who will become a leader in a group (Crockett, 1955). Relevant studies commonly involve leaderless group discussions (Hogan et al., 1994). Thus, no officially appointed leader exists within these groups (Ensari, Riggio, Christian, & Carslaw, 2011). A vast amount of studies within research on emergence of leadership employs the following paradigm. First, participants are measured with regard to a set of traits that function as potential predictors of who is going to emerge as a leader (Guastello, 2007). Second, participants work together in a general problem-solving session or informal discussion. These discussions lead to the exchange of ideas, group interaction, and, importantly, opportunities to take the role of a leader (Ensari et al., 2011). Usually, the outcome of this group interaction is studied, that is, who emerged as a leader (Acton, Foti, Lord, & Gladfelter, 2019). Leadership emergence is usually assessed using questionnaires, asking the group members for their perception about who emerged as a leader in a group, rather than using trained observer ratings (Acton et al., 2019; Ensari et al., 2011). Most research on leadership emergence has used cross-sectional designs, in which the measures for traits and leadership emergence are assessed roughly at the same time (but see,

e.g., Atwater, Dionne, Avolio, Camobreco, & Lau, 1999 for an exception). Additionally, many studies within this field of research assess emergence of leadership via peer ratings, for example in officer training programs in military settings (e.g. Paunonen, Lönnqvist, Verkasalo, Leikas, & Nissinen, 2006), leadership ratings, selections by superiors or attained ranks in military settings (Rueb, Erskine, & Foti, 2008), or ratings by employees in business contexts (e.g., Marinova, Moon, & Kamdar, 2013). Moreover, studies investigating the characteristics of successful candidates in political elections represent a different research field but provide useful insights for investigating emergence of leadership (e.g., Todorov, Mandisodza, Goren, & Hall, 2005).

To exemplify the described leaderless group discussion paradigm, that we focus on in the present study, we describe a study on the effect of narcissism on emergence of leadership (Study 2 in Brunell et al., 2008). In this study, psychology students initially completed several personality questionnaires to measure their scores on the investigated constructs. Subsequently, participants worked in groups of four on a problem-solving task. In this task, they were asked to imagine that they experience a shipwreck and gathered 15 salvageable items from the boat. The participants had the task to rank these 15 items according to their importance for survival. At first, participants individually ranked these items. Subsequently, they were supposed to compare their solutions among each other and reach a collective ranking. Following the discussion, participants were asked to complete a questionnaire with six items asking them to evaluate the extent to which each group member was a leader in the group discussion. An example item was “Group member #1 assumed a leadership role in the group”, and participants were asked to provide a rating for each group member and themselves on 7-point Likert scales from 1 (*not at all*) to 7 (*very much*). For each group member, these scores were used to calculate an average group rating as a leader, as indicated by the other three group members. Additionally, a self-rating as a leader for each participant

was calculated by averaging the ratings a participant gave herself on the six items. Furthermore, participants indicated on a single item the degree they desired to be the leader of the group, using the same Likert scales as for the other rating items. The results of this study show that narcissism is a predictor for all assessed measures of leadership emergence.

As a first overview over predictors for leadership emergence, we start by summarizing meta-analytical findings. Although the focus on identifying the factors that decide who will become a leader in a group started early within research on leadership, interest in it diminished after influential reviews, namely by Stogdill (1948) and Mann (1959), concluded that there are no reliable inter-individual differences associated with emergence of leadership. Researchers regained interest in this topic when re-analyses of existing data challenged that conclusion. Applying a meta-analysis approach to the data used in Mann's (1959) review, Lord and colleagues (1986) indeed found intelligence, masculinity, and dominance to be significant predictors of leadership emergence. As pointed out by House and Aditya (1997), a re-analysis of studies originally analyzed by Stogdill (1948) revealed substantial relationships between intelligence, prosocial assertiveness, self-confidence, energy-activity, and task-knowledge, when studies involving children or adolescents were excluded. More recently, a very influential meta-analysis on personality and leadership showed that four of the five fundamental personality traits (big five) could predict leadership emergence, namely extraversion, openness to experience, conscientiousness, and neuroticism (Judge et al., 2002). In the following, we will provide a more detailed description of the research findings on already identified predictors of leadership emergence.

Overview on Predictors of Emergence of Leadership

Within previous research on emergence of leadership, a broad range of predictors was examined. Acton and colleagues (2019) provided a comprehensive overview of individual characteristics which have an impact on leadership emergence (see Acton et al. [2019], Table

1). We will use this review to structure our overview of the identified predictors of leadership emergence.[1] While doing this, we will also extend that overview with additional predictors of leadership emergence. In the following overview, we will present research findings on the predictors emotional competency, emotional intelligence, and emotional recognition; gender; intelligence, cognitive ability, and knowledge; masculinity, femininity, androgyny (gender role); motivation to lead; personality; physical features; race, self-efficacy; self-esteem/confidence; and self-monitoring.

Emotional competency, emotional intelligence, and emotional recognition. Emotional competency, emotional intelligence, and emotional recognition have been shown to be predictors of leadership emergence. Emotional intelligence, as a prominent example within this category of emotional skills, is considered as a set of abilities referring to processing emotions and emotional information (Côté, Lopes, Salovey, & Miners, 2010). In general, emotional competency, intelligence, and recognition are identified as predictors for leadership emergence, and which particular emotional skill is needed depends on the group requirements (Acton et al., 2019). Although it is claimed that emotional skills play an important role as predictors for emergence of leadership, there are methodological issues in measuring these constructs, that are, for example, questioning the validity of emotional intelligence (see Côté et al., 2010).

Gender. As summarized by Acton and colleagues (2019), previous research showed that men tend to emerge as task-oriented leaders, whereas women tend to emerge as social-oriented leaders. Moreover, men are more likely to emerge in groups working together in the short-term and involving less complex interaction.

---

[1] In this overview, leader behaviors are included as well. Because the present research does not focus on behaviors as predictors for emergence of leadership, we will not consider them in the following.

Intelligence, cognitive ability, and knowledge. Previous research showed that intelligence, cognitive ability, and knowledge all are important determinants of leadership emergence (Acton et al., 2019). As an example of this category, we will focus on the relationship between intelligence and leadership emergence. Intelligence represents the general ability to handle abstractions, new situations, and learning. It includes the ability to adjust to new circumstances and to solve any kind of problem (Rueb et al., 2008). A general link between intelligence and leadership emergence is commonly found (Ensari et al., 2011; Foti & Hauenstein, 2007; Gershenoff & Foti, 2003; Kickul & Neuman, 2000; Rueb et al., 2008; J. A. Smith & Foti, 1998; Taggar et al., 1999). Accordingly, more intelligent individuals are more likely to emerge as leaders (Acton et al., 2019). Intelligence was already revealed as a significant predictor (r = .51) of leadership emergence in the comprehensive meta-analysis by Lord et al. (1986) and was considered as a key characteristic in predicting leadership perceptions early on. A more recent meta-analysis (Judge, Colbert, & Ilies, 2004a) confirmed that intelligence is related to leadership emergence, thereby revealing that the relation between perceptual measurements (e.g., ratings of group members) and leadership emergence is higher than with objective measurement of intelligence (e.g., an intelligence test). Intelligence might cause a person to be perceived as a leader, tapping into a person's implicit leadership theory (Judge et al., 2004a). Rubin et al. (2002) found that both objective intelligence measures predicted leadership emergence, but also that perceived intellectual competence was related to leadership emergence. Furthermore, perceived intelligence mediated the effect of actual intelligence on leadership emergence.

Masculinity/femininity/androgyny (gender role). Masculinity and femininity are defined as two independent facets of personality and follow the rationale that individuals differ in the degree they are masculine (high in masculinity and low in femininity), feminine (low in masculinity and high in femininity), androgynous (high in masculinity and

femininity) or undifferentiated (low in both dimensions; Moss & Kent, 1996). Masculinity was already found to be a predictor of leadership emergence in the first meta-analysis on factors predicting leadership emergence (Lord et al., 1986), and a more recent meta-analysis showed that masculinity was a significant predictor for leadership emergence, whereas femininity was not (Ensari et al., 2011). Previous research showed that masculine and androgynous individuals are more likely to emerge as leaders than feminine individuals. In general, masculinity serves as a common schematic category for leaders (Acton et al., 2019).

Motivation to lead. Motivation to lead (MTL) is the motivation of a person to assume leadership roles because of self-inclination and preference (Hong, Catano, & Liao, 2011). In other words, this motivation describes the motivation to lead simply because a person enjoys leading (Suessenbach, Loughnan, Schönbrodt, & Moore, 2018). Acton and colleagues (2019) summarize that individuals having a high motivation to lead tend to emerge as leaders. Moreover, it is considered that these individuals are more resilient in experiencing setbacks while pursuing a leadership role (Acton et al., 2019).

Personality. The five-factor model of personality (the big five) represents a comprehensive model of personality (Judge et al., 2002). As the name implies, it consists of five facets, namely agreeableness, conscientiousness, extraversion, neuroticism, and openness to experience. The arguably most influential paper on the relationship between the big five facets and leadership emergence is the comprehensive meta-analysis by Judge and colleagues (2002). In general, the authors find a multiple correlation of $R = .53$ between the big five and leadership emergence, indicating that personality in general and the big five traits specifically are promising candidates when investigating the inter-individual differences of a leader.

Agreeableness. Agreeableness captures the tendency to trust and care, as well as being compliant and gentle (Judge et al., 2002). Results from two meta-analyses (Ensari et al., 2011; Judge et al., 2002) did not find significant relationships between agreeableness and

leadership emergence. Agreeable individuals might be less likely to emerge as a leader in a group because they are considered to be passive and compliant (Judge et al., 2002).

Conscientiousness. Conscientiousness combines two dimensions, namely, achievement and dependability (Judge et al., 2002). Previous research identified conscientiousness as a predictor for leadership emergence (Ensari et al., 2011; Judge et al., 2002). It is plausible to assume that the achievement facet of conscientiousness is the driving factor behind this relationship. Conscientious individuals aim for achievement and could fulfill this tendency by emerging as a leader. Moreover, conscientious individuals engage in organizing activities that could help them to emerge as leaders rather quickly (Judge et al., 2002).

Extraversion. Extraversion describes the tendency to be social, assertive, and the ability to experience positive affects (e.g., energy). This construct consists of the two facets sociability and dominance (Judge et al., 2002). Previous research consistently found extraversion to be one of the strongest and most important predictors for leadership emergence (Ensari et al., 2011; Judge et al., 2002). Extraversion is related to being perceived as leader-like (Judge et al., 2002). Social and dominant individuals are expected to assert themselves in group situations, and therefore, extraverted individuals are likely to emerge as leaders (Judge et al., 2002).

Neuroticism. Neuroticism is the tendency to experience negative affects (e.g., anxiety) and to display emotional maladjustment (Judge et al., 2002). Meta-analytically, neuroticism appears to be significantly and negatively correlated with leadership emergence. However, it failed to reach significance in predicting leadership emergence together with the other big five traits in a multiple regression analysis, probably because of its intercorrelation with the other big five facets (Judge et al., 2002). In a more recent meta-analysis, neuroticism appeared to be negatively related to leadership emergence but failed to reach significance

(mean Fisher's Z = -.08, p = 0.06; Ensari et al., 2011). Individuals high in neuroticism are less likely to be perceived as leaders (Hogan et al., 1994). Moreover, self-esteem is considered as a predictor for leadership emergence and might serve as a proxy for low neuroticism (Judge et al., 2002). In sum, the evidence for the relation between neuroticism and leadership emergence is not as clear as for the other big five traits.

Openness to experience. Openness to experience describes the tendency of being imaginative, unconventional, and autonomous. Meta-analytically, openness to experience was shown to be a predictor of leadership emergence (Ensari et al., 2011; Judge et al., 2002). Openness to experience was found to correlate as strongly as extraversion with leadership in business settings (Judge et al., 2002).

Physical features. In general, physical features are considered as predictive for leadership emergence. Among these features, Acton and colleagues (2019) specifically mention physical fitness, height, and attractiveness as predictors of leadership emergence. In the following, we will shortly describe why physical features might be related to leadership emergence before describing research findings on the mentioned physical features and leadership emergence.

These features are believed to be predictive of leadership emergence because they are considered as having evolutionary benefits (Acton et al., 2019). Although leaders in modern times usually work in offices and mostly perform cognitive tasks, there are findings linking, to name just one example, body height to salaries (Judge & Cable, 2004). This relationship might have evolutionary roots. Considering the animal kingdom, where bigger animals appear more dangerous, animals use height as a cue to infer power and strength in fight-or-flight decisions. In that sense, height is an indicator of power and therefore asks for respect (Judge & Cable, 2004). Evolutionary leadership theory (ELT) considers leadership in the context of evolutionary psychology and provides a framework to explain the relationship

between physical features and leadership. Evolutionary psychology assumes that evolution, driven by natural selection, shaped our human minds, just as it shaped our bodies (Van Vugt & Ronay, 2014). Furthermore, evolutionary psychology assumes that there are psychological mechanisms embedded in our minds today that originally helped solving problems critical for survival and reproduction in earlier stages of human development. These assumed mechanisms work like "if-then" decision rules. For example, the rules "follow a physically fit individual" or "follow a tall individual" might have evolved because it proved to be a successful strategy in terms of providing reproductive benefits. These rules do not necessarily work consciously (Van Vugt & Ronay, 2014). Evolutionary leadership theory calls the phenomenon that characteristics of a leader, that might have been important in ancient times, but still influence our leadership perceptions a "mismatch" (Ahuja & Van Vugt, 2010). Human society changed a lot in the last 10000 years, moving from hunter-gatherer societies to contemporary settled communities. From an evolutionary point of view, this is only a short period of time, and we might not have changed so much. As a result, we might still select leaders according to their physical and psychological traits that were desirable in former times but do no longer provide benefits in our modern societies (Ahuja & Van Vugt, 2010). ELT suggests that we humans have internalized a cognitive ancestral prototype of a leader. This prototype evolved and served as a template of how a leader is supposed to be like (Ahuja & Van Vugt, 2010).

Body height. Body height was reported to be a predictor for career success (Judge & Cable, 2004). Body height is an important factor for candidates in US presidential elections, whereby taller presidents were not more likely to win the actual election but were more likely to be re-elected and received more popular votes (Stulp, Buunk, Verhulst, & Pollet, 2013). Previous research showed that tall people are more perceived as leader-like. Body height belongs to desirable physical features linked to evolutionary fitness, and therefore, tall

persons might be automatically and unconsciously perceived as leaders (Blaker et al., 2013). In a meta-analysis, Judge and Cable (2004) showed that body height is significantly correlated with leadership emergence. Moreover, the authors also found a significant correlation between height and self-esteem, which was significantly higher than the observed correlation between body height and leadership emergence. The authors claim that tall individuals are treated with respect on a regular basis and may, therefore, become more self-confident. In a similar way, they might have made the experience that they were successful in the past and therefore, may be more self-confident and more confident in their competence and abilities.

Physical fitness. Physical fitness has been identified as a predictor for leadership emergence, whereby it was conceptualized in different ways that we describe in the following. Physical fitness, as measured via a physical fitness test, was predictive for the level of rank a (male) cadet in a military college attained (Atwater et al., 1999). Similar findings between physical fitness and leadership emergence were found in a study by Rice, Yoder, Adams, Priest, & Prince (1984). Using the waist-to-hip-ratio as an indicator for physical fitness, a study by Campbell et al. (2002) found physical fitness to be predictive of leadership emergence in males during leaderless group discussions in two studies. The authors argue that one of the reasons why the waist-to-hip-ratio serves as a cue for health and underlying fitness is because it is a reliable measure of body fat distribution. Moreover, it predicts a variety of health risk factors. For men, desirable waist-to-hip-ratios range from .90 to .95, and this range is the one associated with the best long-term health outcomes (Campbell, Simpson, Stewart, & Manning, 2002). As shown by Atwater et al. (1999), physical fitness was also related to self-esteem.

Perceived attractiveness. Physical attractiveness is defined as the extent to which it is pleasing to watch a person (Patzer, 1985). Per definition, an attractive person is considered to

be beautiful and above average good-looking (Umberson & Hughes, 1987). Previous research on the effects of attractiveness on leadership emergence mainly focused on presenting the persons' faces. The face seems to be an important source of information, and persons tend to make inferences on a person's personal dispositions from the face (Hassin & Trope, 2000). Evidence on attractiveness and emergence of leadership revealed that judged by their yearbook photo, more attractive and more mature looking students were more likely to be recognized as leaders (Cherulnik, Turns, & Wilderman, 1990). A recent study (Gruber, Veidt, & Ortner, 2018) showed that women with higher facial attractiveness had higher ascribed social competence which significantly predicted the number of leader nominations she received from her group members in an all-female group, working on assessment center tasks. In this study, the relation between facial attractiveness and leadership emergence failed to reach significance. Attractiveness is considered as a significant predictor for political elections (e.g., Antonakis & Dalgas, 2009; Antonakis & Eubanks, 2017; Todorov et al., 2005). A study on that topic found that more attractive candidates got more votes than less attractive candidates and won elections more often (Efrain & Patterson, 1974). A study investigating the effects of beauty, perceived competence, and perceived trustworthiness showed that beauty was strongly related to election success in a Finnish election (Berggren, Jordahl, & Poutvaara, 2010). Attractive persons might have such an advantage because compared to less attractive persons, they are ascribed positive characteristics, which is described as a generic halo effect (Van Vugt & Grabo, 2015). For example, attractive persons are perceived to be more socially competent and successful compared to less attractive persons (Langlois et al., 2000). From an evolutionary point of view, attractiveness might be considered as an indicator of health and reproductive fitness (Van Vugt & Grabo, 2015; Weeden & Sabini, 2005). Moreover, previous research showed that perceived attractiveness is related to a person's core self-evaluations. As a person's core self-evaluations include her

basic evaluations of herself and her success and control over her life, this construct is related, among others, to a person's self-esteem and general self-efficacy (Judge, Hurst, & Simon, 2009). Therefore, these results indicate that persons that are more attractive might have higher levels of self-confidence and general self-efficacy, which in turn might increase the probability of them emerging as leaders in a group.

Race. As described by Acton and colleagues (2019), race seems to have an effect on leadership emergence in that the fit between race and occupation predicts leadership emergence, the congruence between race and task type increases the perception of leadership, and leader prototypes differ on race, driving leader perceptions (Acton et al., 2019).

Self-efficacy. Self-efficacy, or general self-efficacy, describes the expectation of being able to properly execute behaviors to reach a desired outcome (J. E. Smith et al., 2015). More specifically, it includes the expectation to perform successfully in diverse, challenging situations, even when achieving the outcome calls for effort and perseverance (Amos & Klimoski, 2014). The link between general self-efficacy and leadership emergence has been shown in several studies (Foti & Hauenstein, 2007; Serban et al., 2015; J. A. Smith & Foti, 1998). It is assumed that individuals with high self-efficacy are more confident and therefore, more likely to attain leadership roles (Acton et al., 2019). Individuals with high general self-efficacy might have more self-confidence because they experience successes in their lives and show more effort and perseverance than individuals low in self-efficacy. The resulting self-confidence might even influence their expectations of self-efficacy (J. A. Smith & Foti, 1998). Meta-analytically, general self-efficacy was shown to be a significant predictor of leadership emergence (Ensari et al., 2011).

Self-esteem/confidence. Self-esteem and confidence are considered as important for leadership emergence (Acton et al., 2019). In the following, we will focus on self-esteem. Self-esteem captures how individuals value themselves (Judge & Cable, 2004) and more

specifically, reflects the positive or negative attitude an individual has regarding herself (Ferring & Filipp, 1996). Self-esteem was identified as a predictor for leadership emergence in military settings (Atwater et al., 1999; Paunonen et al., 2006) and was identified as a significant predictor for leadership emergence in a meta-analysis (Ensari et al., 2011). Amos and Klimoski (2014) proposed that to predict whether an individual takes risky actions, like making a leadership attempt, an individual's view on herself is a critical factor.

Self-monitoring. Self-monitoring denotes an individual's skill to observe and control her behavior according to the current situation (Snyder & Gangestad, 1986). High self-monitors are sensitive to the given situation and interpersonal cues regarding their social behavior. They use these cues for regulating and controlling their verbal and nonverbal self-presentation. In contrast, low self-monitors are not that aware of social information and the appropriate self-presentation in a situation. They do not seem to adjust their behavior accordingly (Berkowitz, 1979). Self-monitoring is considered as a strong (Rubin, Bartels, & Bommer, 2002) and consistently found predictor for leadership emergence (Eby, Cader, & Noble, 2003). The link between self-monitoring and emergence of leadership is well established (Eby et al., 2003; Ellis, 1988; Foti & Hauenstein, 2007; but see Rueb et al., 2008; Zaccaro, Foti, & Kenny, 1991). It was found that high self-monitors were more likely to be perceived as leaders when they also score high on intelligence, general self-efficacy, and dominance (Foti & Hauenstein, 2007). The relation between self-monitoring and leadership emergence was found to be mediated by the perceived intellectual competence of an individual (Rubin et al., 2002). In a study using both perceptual and behavioral measurements of leadership emergence, it was shown that high self-monitors were more likely to be perceived and nominated as a leader by their group members than low self-monitors. Also, subjects with strong self-monitoring tendencies showed more task-oriented leader behaviors. High self-monitors did not show more relationship-oriented leader behavior than low self-

monitors (Eby et al., 2003). Interestingly, there was only a partial congruence of the results regarding perceptual and behavioral measurements of leadership emergence in the cited research. This might indicate that, while there is an overlap in what perceptual and behavioral measures of leadership emergence capture, that they are also affected in different ways or by different influencing factors. Eby and colleagues (2003) suggest this might be a consequence of implicit leadership theories influencing participants' perceptions of leadership emergence. We will take a closer look at this idea and at implicit leadership theories in a later section.

Additionally identified predictors for leadership emergence. Extending the overview on predictors for emergence of leadership by Acton and colleagues (2019), we will present additional predictors and their evidence towards leadership emergence in the following, namely perceived trustworthiness, perceived competence, dominance, and narcissism. Some scholars consider dominance and narcissism as sub-facets of extraversion, but we aim at giving an overview of these narrow personality traits and their relation to leadership emergence, too. Evidence for perceived trustworthiness and perceived competence as predictors for leadership emergence mostly comes from research on election studies. Therefore, they might not have been listed in the review by Acton and colleagues (2019).

Perceived trustworthiness. Trustworthiness is defined as the willingness to act favorably towards other persons when there are explicit or implicit demands or expectations of action (Ben-Ner & Halldorsson, 2010). Trustworthiness also includes refraining from exploiting other persons' weaknesses (Barney & Hansen, 1994). According to evolutionary leadership theory, trustworthiness might be a relevant characteristic of a potential leader. For example, trustworthiness might be related to leadership because a decision rule stating to follow an individual that seems trustworthy might have provided evolutionary benefits (Van Vugt & Ronay, 2014). A study by Little, Roberts, Jones, & Debruine (2012) manipulated candidates' faces in a hypothetical election scenario. Participants were asked to elect a leader

for their country and preferred more trustworthy faces compared to less trustworthy faces. In a subsequent study, the authors showed that more trustworthy faces were voted for within peacetime compared to a wartime scenario. In contrast, two studies did not find significant effects of trustworthiness on election success (Berggren et al., 2010; Todorov et al., 2005). Summing up, the relationship between trustworthiness and election success is less established. Evidence from the business context suggests an important role of trustworthiness for leadership positions. A study by Linke, Saribay, & Kleisner (2016) found that their position in an organizational hierarchy correlated with perceived trustworthiness of male managers' faces. Higher ratings of perceived trustworthiness were associated with a higher number of subordinates and a smaller number of superiors. The position in the hierarchy did not correlate with perceived attractiveness or dominance. Another study indicates that perceived trustworthiness might be important when a CEO is supposed to be selected after financial misconduct in a company (Gomulya, Wong, Ormiston, & Boeker, 2017).

Perceived competence. Competence refers to an individual's abilities, proficiencies, or dispositions to learn or do something successfully or to reach a goal (Weinert, 1999). There is substantial evidence suggesting that election results can be predicted by perceived competence. Todorov, Mandisodza, Goren, and Hall (2005) were able to predict which of two candidates will win elections for the U.S. Congress based on competence ratings of the candidates' pictures. The candidate rated as more competent won in 71.6% of the Senate races and in 66.8% of the House races. Neither attractiveness nor trustworthiness were significant predictors in this study. In a study predicting the election results of the 2002 French parliamentary election, Antonakis and Dalgas (2009) could also show that perceived competence serves as a predictor for election success. This study is remarkable because, in addition to adults, children were asked to select one of the candidates as the captain of a boat. The results revealed that adults and children were equally successful in predicting an election

result. This suggests that adult voters and children use similar cues to evaluate a candidate's competence from his face when selecting a leader. In a different study investigating elections, perceived competence, as well as perceived attractiveness and perceived trustworthiness, did not serve as a significant predictor for election success (Berggren et al., 2010).

Dominance. Dominance is defined as the disposition to try to control the environment, to influence others, to speak out one's opinion, and to be able to take a leadership role spontaneously, while at the same time liking this role. Dominance includes being decisive and assertive, but also influential and persuasive (Jackson, 1974; Stumpf, Angleitner, Wieck, Jackson, & Beloch-Till, 1985). Dominance was already identified as a predictor in the meta-analysis of Lord et al. (1986) and was consistently associated with leadership emergence since then (Ensari et al., 2011; Foti & Hauenstein, 2007; Rueb et al., 2008; J. A. Smith & Foti, 1998). An important question is how dominant individuals attain leadership positions in groups. Anderson and Kilduff (2009) assume that besides behaving forcefully, dominant individuals might also possess high task-related abilities and general leadership skills. In a study investigating their assumption, the authors found that dominant individuals were rated as more competent and that these ratings partially mediated the effect of dominance on influence in groups.

Narcissism. Narcissism, seen as a personality trait rather than in terms of a narcissistic personality disorder, consists of several facets. These include positive and inflated self-views, maintaining these through pervasive behaviors, even if it hurts others, and interpersonal relationships missing warmth and intimacy (Brunell et al., 2008). Previous research found narcissism to be a predictor for leadership emergence (Brunell et al., 2008; Grijalva, Harms, Newman, Gaddis, & Fraley, 2015; Nevicka, De Hoogh, Van Vianen, Beersma, & McIlwain, 2011; Paunonen et al., 2006). As argued by Brunell and colleagues (2008), there are several reasons why narcissism might be related to leadership emergence.

Narcissists seem to desire social status and power, including leadership roles, and they are skillful at initiating relationships with persons they are not acquainted with. Moreover, in group situations without a formal leader, narcissists are expected to use this opportunity for self-enhancement (Brunell et al., 2008). Narcissists might be perceived as leaders in a group because of their tendency to show overconfidence. This increased level of confidence might make it more likely that narcissists speak up in a group situation (Brunell et al., 2008). A study by Brunell and colleagues (2008) showed that participants scoring high on narcissism were perceived as the leaders in their groups, but they were not successful in enforcing their opinion.

This finding indicates that emergent leaders are not necessarily successful in their leadership attempts. Therefore, it might be an interesting research question to differentiate a leadership attempt from its success. Following that idea, we present the goals of the present research.

The Present Research

Our present research has four goals. First, we want to differentiate emergence of leadership into two distinct concepts, namely leadership initiative and the success of leadership initiative. As a second goal, we want to introduce a behavioral measurement approach to emergence of leadership. Third, we aim to examine how previously neglected predictors influence emergence of leadership. For our fourth and last goal, we take a closer look at the effects of gender on leadership initiative by systematically manipulating the gender composition of the examined ad-hoc groups in our studies.

First goal. The first goal we identified considering previous research is that many researchers regarded and, hence, operationalized emergence of leadership as a global construct with a strong focus on leaders. Although effective leadership does, by definition, require both leaders and followers, the role of the followers has been mostly neglected so far

(Uhl-Bien, Riggio, Lowe, & Carsten, 2014; e.g., Van Vugt, 2006). More recently, researchers have begun to consider emergence of leadership as a mutual social influence process and, therefore, included the followers' role. For example, one of the latest definitions of emergence of leadership states that, during the process of emergent leadership, patterns of leader and follower interactions are built (Acton et al., 2019). Within the literature, there are first ideas about differentiating emergence of leadership and taking the role of followers into account (e.g., Emery, Calvard, & Pierce, 2013; Van Vugt, 2006). Based on these ideas, we want to introduce a strict differentiation between leadership initiative and its success in attracting followers.

As illustrated, the role of followers can be crucial because their behavior constitutes the emergent leader's success. Therefore, it is reasonable to investigate whether an emerging leader is successful in attracting followers. For our research, we aim to predict separately who is showing leadership initiative and who is successful in attracting followers. Inter-individual differences do not necessarily have to predict both components of leadership emergence; there might be factors that predict only one component of emergence of leadership. Conceptually, both components are not fully independent of each other. If a person does not initiate leadership, she cannot attract followers. Similarly, initiating leadership does not necessarily lead to success. A narcissistic individual in a group might show leadership initiative, but the other group members do not have to follow that attempt.

For the separation of leadership emergence into the proposed components, we drew inspiration from research on emergence of leadership in non-human primates. This conceptual differentiation is already used in non-human primate research. As described by King, Douglas, Huchard, Isaac, & Cowlishaw (2008), animals living in groups have the need to coordinate their actions in order to maintain group cohesion. In general, these animals can make decisions in two ways. In a consensus decision, all group members contribute

democratically to a decision. In a despotic decision, considered as the extreme opposite, a single animal makes the decision. This animal is called the "leader", with other animals accepting that decision, who are called "followers" (King et al., 2008). King et al. (2008) point out that these despotic decisions are common among group-living vertebrates, including humans. In this way, differentiating emergence of leadership in a leader and a follower component seems a reasonable approach. Moreover, we drew inspiration for our research paradigm from research on collective group movements in non-human primates. We might consider the successful initiation of a collective movement of a group as leadership (Fischer & Zinner, 2011). For example, chacma baboons initiate group movements according to a specific pattern of behavior (Stueckle & Zinner, 2008). As described by Fischer and Zinner (2011), these baboons rest at their sleeping site before starting their daily march. To initiate group movement, one individual, which is called "the initiator", moves away from her group. Her attempt was successful if others followed her; if the others stayed behind, the attempt was not successful. The study showed that there was a critical mass of followers needed to get the group moving. If too few individuals followed the attempt, the initiator went back to the group, until either the same or another individual made another attempt to initiate movement.

To sum up, we aim at splitting the emergence of leadership construct into two components: a) "leadership initiative": a person initiates leadership behavior and b) "success of leadership initiative": the other group members follow a person who initiated leadership behavior. It is important to note that we are focusing on the emerging leader's individual characteristics in determining her success in attracting followers. Investigating the inter-individual differences leading a person to follow a leadership attempt is a different research question that we do not address in the present research.

Second goal. As the second goal of our studies, we wanted to test whether a broad range of previously identified predictors for the perceived emergence of leadership could also predict emergence of leadership on a behavioral level. This research goal is twofold: First, emergence of leadership is traditionally measured as the degree to which other group members perceive a person as a leader. This perceptual measurement can be problematic because it might be confounded with raters' implicit leadership theories. Participants in a group discussion might perceive another group member as a leader because she appears to be "leader-like" and matches their stereotype of a leader even if she is not actually leading the group. Second, by predicting emergence of leadership with all our predictors simultaneously, we investigate whether some of them might be dispensable because of redundancies.

The term "implicit leadership theories (ILT)" denotes the implicit and naive conceptualizations individuals have of leaders. They reflect what is generally expected from leaders concerning their traits or attributes (Offermann & Coats, 2018). In other words, individuals have unconscious, implicit leadership theories, which help them to distinguish leaders from non-leaders (Shondrick & Lord, 2010; Shondrick, Dinh, & Lord, 2010). Conceptually, ILTs derive from leader categorization theory and represent a simple type of category system (Lord, Foti, & De Vader, 1984). As described by Offerman et al. (1994), leadership is a hierarchically organized cognitive category. When an individual, called "the perceiver" in the following, perceives a stimulus person, she makes the basic distinction whether that person is a leader or non-leader. There should be a few characteristics that apply to all leaders and little overlap between leaders and non-leaders. Subsequently, the perceiver classifies the stimulus person according to one of her leader prototypes. Stated differently, a person is perceived or accepted as a leader if she sufficiently matches the leader prototype of the perceiver. ILTs are considered to be formed and adjusted over time using a person's

experience with actual leaders or descriptions of leaders (Offermann, Kennedy, & Wirtz, 1994).

In the context of the present research, ILTs play an important role. Since previous studies assessed emergence of leadership almost exclusively on a perceptual level, they most likely collected subjective ratings of emergent leaders. Such ratings, at least partially, might reflect raters' implicit leadership theories rather than who objectively took the lead in a group (Eby et al., 2003). Imagine a narcissistic person who appears very intelligent and superior working in a group, but who is actually only using the stage to impress the other group members without really leading the group (which would be indicated behaviorally by successfully initiating leadership). Matching the other group members' implicit leadership theories, she still might be perceived as the group's leader.

There is a large body of research indicating which inter-individual differences predict who is perceived as a leader. However, the question of whether the group members' perceptions are in line with who is actually leading the group remains unanswered. Because of the potential confound between the group members' implicit leadership theories and their perceptions of objective leadership emergence, our goal is to introduce a behavioral measurement of emergence of leadership.

Behavioral measures of leadership emergence have been rarely used in previous studies (cf. Brunell et al., 2008; Cronshaw & Ellis, 1991; Eby et al., 2003; Foti & Hauenstein, 2007). Hence, we have only very limited evidence on whether perceptual and behavioral measurements of leadership emergence correspond to each other. Evidence from the few studies using behavioral measurements of leadership emergence shows only partial (Eby et al., 2003) or no support for the accordance of these two types of measures (Brunell et al., 2008). The study by Eby and colleagues (2003) is a good example of the behavioral assessment of leadership emergence. The authors showed that participants scoring high on

self-monitoring engaged in significantly more task-oriented leader behaviors than participants low on self-monitoring. Task-oriented leader behaviors included planning and organizing within the group discussion. Trained raters scored the participants' behavior during the group interaction according to a behaviorally based inventory of leader behaviors, that was developed by the same authors, namely the Observational Inventory of Leader Behaviors (OILB). According to the inventory, a component of task-oriented leader behavior is providing structure to the situation, for example, by saying, "Why don't we start by telling what we thought and why?" (Eby et al., 2003, p. 1466). In this study, leadership emergence was also assessed using the other group members' rated perceptions of the emergent leader. As we mentioned earlier, the authors found that there was only a partial congruence between their perceptual and behavioral measurements of leadership emergence. Hence, perceptional and behavioral measurements of leadership emergence do not necessarily assess the same aspects of the phenomenon. In a study by Brunell and colleagues (2008), participants scoring high on narcissism were perceived as leaders in their groups, but they were not able to enforce their opinion. More recent approaches identified nonverbal features for the identification of emergent leaders in small groups, like body pose (Beyan, Katsageorgiou, & Murino, 2017), head pose (Beyan et al., 2016), and audio-visual cues (Sanchez-Cortes, Aran, Jayagopi, Schmid Mast, & Gatica-Perez, 2013; Sanchez-Cortes, Aran, Mast, & Gatica-Perez, 2012). In sum, it is unclear which of the already identified predictors for emergence of leadership are still predictive when a behavioral measurement of leadership emergence is applied instead of a perceptual measurement.

Third goal. Third, we aimed at investigating the effect of predictors that up to now were mostly neglected in the research on emergence of leadership. Thereby, we included newly identified predictors and used approaches in measuring constructs that were not yet covered by previous research. We will present these purposes in the following.

Testosterone. As a physical feature that was neglected by previous research on leadership emergence, we aimed at investigating the effect of testosterone on leadership emergence. Testosterone is a steroid hormone produced in both men and women. Basal testosterone levels (i.e., baseline testosterone levels) were shown to be stable over time (Liening, Stanton, Saini, & Schultheiss, 2010) so that testosterone is considered as a stable inter-individual difference in research. Testosterone has an influence on social behavior for both sexes, even though men have higher levels of testosterone than women (see Liening & Josephs, 2010, for an overview).

There are several findings suggesting a link between testosterone and leadership. Dominance is considered to be the primary psychological effect of high testosterone levels (Liening & Josephs, 2010). Dominant individuals are motivated to attain high status, and they also succeed in achieving that high status (Anderson & Kilduff, 2009). Testosterone might be associated with leadership because high testosterone levels are linked to high social status, and leaders are ascribed to have higher social status than non-leaders (Van der Meij, Schaveling, & Van Vugt, 2016). Moreover, leadership is often attained through dominance, and, as already described, dominance is related to testosterone (Van der Meij et al., 2016). Additionally, testosterone was as well linked to aggression in humans (Archer, 2006). This might be considered a possibly detrimental effect of testosterone on the success in attracting followers. While a meta-analysis on leadership and testosterone did not find a difference in testosterone levels for managers compared to non-managers, testosterone still might be relevant for attaining leadership positions (Van der Meij et al., 2016). To sum up, there is little evidence telling us whether testosterone is predicting who is emerging as a leader in leaderless groups.

Risk-taking tendency. Although the act of emerging as a leader involves taking risks (Amos & Klimoski, 2014), the effect of an individual's tendency to take risks on emergence

of leadership has, so far, not been investigated. Risk-taking is defined as the tendency to engage in behaviors, even when these might lead to negative outcomes (Boyer, 2006; Dekkers et al., 2019). Inter-individual differences in preferring risks affect leadership (Bass & Bass, 2008). An individual's attitude towards risk might be a critical factor when it comes to decisions that are made under uncertainty (i.e., the correct solution to a problem is unclear). For example, the risk of a decision is evident for top managers who face competition (Bass & Bass, 2008). In many situations, the correct decision is not obvious, and leaders have to take risks in deciding for a course of action. To the best of our knowledge, the effect of risk-taking on leadership emergence had not been investigated, yet.

Subjective competence. Subjective competence refers to an individual's subjective assessment of the performance-related abilities and skills that are required to solve the task at hand (Weinert, 1999). It is highly plausible that an individual's subjective competence in solving a given task has an effect on leadership emergence. It is reasonable to assume that persons who assess their competence in a given task as high might tend to engage in leadership behavior, and therefore, might show leadership attempts. Although the effect of subjective competence was neglected in previous research, we believe that subjective competence differs between individuals and represents a reasonable addition to traditionally investigated predictors for leadership emergence.

Measuring motives: Explicit and implicit power motives. Within the present research, we also aimed at investigating the effect of power motives on leadership emergence. We investigated different types of measurements (explicit and implicit measurement), and also examined newly developed sub-facets of the explicit power motive, namely, the dominance, prestige, and leadership scales (Suessenbach et al., 2018). In the following, we will present these different approaches with regard to the emergence of

leadership, starting with a short general background on measuring explicit and implicit motives.

People do not only differ in terms of personality but as well in their motives. In general, motives are considered to operate on both an implicit and an explicit level: implicit motives concern nonconscious, affective preferences for certain classes of incentives. In contrast, explicit motives are self-attributed and reflect the self-concept of a person regarding their goals, values, personality attributes, and affective preferences. The difference between explicit and implicit motives becomes clear in their measurements, as well. Since the implicit motives are considered to operate unconsciously, they are measured indirectly, for example using the Picture Story Exercise, whereas explicit motives can be assessed via self-report questionnaires (Schönbrodt & Gerstenberg, 2012). Among the implicit motives, there are three different motives, namely the power motive, the achievement motive, and the affiliation motive. In the following, we will focus on the power motive. The power motive is defined as "a concern for having an impact on others, arousing strong emotions in others, or maintaining reputation and prestige" (Winter, 1988, p. 510).

A meta-analysis on the relation between explicit and implicit motives revealed that the explicit and implicit power motive both do not significantly correlate with each other (Koellner & Schultheiss, 2014). This finding supports the assumption that a desire for power is socially undesirable (McClelland, 1987). Since both measures do not seem to overlap, and therefore, seem to be independent of each other, we included both measures in our study. In this context, the implicit measurement is considered to be unbiased, and the explicit measurement serves as a comparison. We included the implicit power motive in the present research to investigate the effect of an implicit measurement of this motive on leadership emergence. Additionally, we included recently developed sub-facets of the explicit power motive, namely the dominance, prestige, and leadership motive scales (Suessenbach et al.,

2018), that we describe in the following. For exploratory purposes, we assessed a global measure of the explicit power motive. In the following, we will further present the implicit and explicit power motive in more detail.

Implicit power motive. Previous research discovered associations between the implicit power motive (among others) and leadership performance (e.g., Steinmann, Dörr, Schultheiss, & Maier, 2015). There is evidence suggesting that the power motive is related to outcomes associated with power. For example, the power motive predicts the attainment of professions with high power (Winter, 1988). However, to the best of our knowledge, a possible effect of the implicit power motive on emergence of leadership within classical emergence of leadership research has not yet been investigated.

Explicit power motive: Global assessment. As already described, we assessed the explicit power motive for exploratory purposes using a currently developed inventory, the Unified Motive Scale (UMS; Schönbrodt & Gerstenberg, 2012). There is evidence that the explicit power motive is a predictor of leadership emergence (Thomas, Dickson, & Bliese, 2001). Within research on emergence of leadership, the exact assessment of the explicit power motive became blurred with other traits: Self-reported dominance is sometimes considered as a measure for the explicit power motive, although dominance is oftentimes referred to as a trait, rather than a motive (e.g., Foti & Hauenstein, 2007; J. A. Smith & Foti, 1998). In the present research, we consider dominance rather as a trait than a motive. Additionally, we aim at investigating the unique predictive value of single predictors. Therefore, we decided to include the global explicit power motive but decided to use a new inventory in order to assess it. We included the global explicit power motive for exploratory purposes.

Explicit power motive: Sub-facets dominance, prestige, and leadership motive. Instead of focusing on the explicit power motive as a general construct, we decided to use its

recently proposed sub-facets, namely the dominance, prestige, and leadership motive (Suessenbach et al., 2018). These facets further differentiate the explicit power motive with regards to the research on leadership emergence. As already noted, previous research found evidence that a related construct, motivation to lead, is a predictor of leadership emergence. As we were interested in exploring the differential effects of the new subscales of the explicit power motive, we decided to use the dominance, prestige, and leadership motive. To the best of our knowledge, the association between the dominance, prestige, and leadership motive and emergence of leadership has not yet been investigated. In the following, we briefly describe the sub-facets.

The dominance motive is defined as the "desire to coerce others into adhering to one's will" (Suessenbach et al., 2018, p. 8). This might be achieved through dominant behaviors, like verbal aggression or manipulating others (see Suessenbach et al., 2018, for an overview). The prestige motive is defined as the desire to gain respect and admiration from others, especially for a person's skills and knowledge (Suessenbach et al., 2018). The leadership motive is defined as the "desire to take initiative and responsibility in one's group to direct it to a common goal" (Suessenbach et al., 2018, p. 9). This construct essentially captures the core of leadership. The leadership motive was shown to significantly predict the number of self-reported leadership positions an individual has had and correlates significantly with an individual's employment rank (Suessenbach et al., 2018).

Fourth goal. Furthermore, we aim to investigate systematically whether group gender composition influences the effect of inter-individual differences on leadership emergence. Previous research investigated groups consisting of both women and men (e.g., Brunell et al., 2008; Lemoine, Aggarwal, & Steed, 2016), only men (e.g. Campbell et al., 2002), especially in military contexts (e.g., Atwater et al., 1999; Paunonen et al., 2006), or female-only groups (e.g. Gershenoff & Foti, 2003; Gruber et al., 2018). The available evidence so far suggests

that the effect of inter-individual differences on emergence of leadership could be moderated by group gender. For example, extraversion has a stronger impact on emergence of leadership in male groups when the performance is evaluated by an attractive woman (Campbell, Simpson, Stewart, & Manning, 2003). However, the effect of group gender composition was rarely investigated in previous research (but see Lemoine et al., 2016). To the best of our knowledge, group gender composition was not yet experimentally manipulated, although predictors might be of different relevance in, for example, male versus female groups. Up to now, it remains unclear whether group gender composition differentially influences leadership initiative and its success, and we aim at investigating that research question. For exploratory purposes, the present research aims at investigating whether the effects of our predictors are robust across different group gender compositions, namely mixed-gender, male, and female groups.

As the first step in order to implement these four research goals, we make a selection of predictors that we aim to use. We will explain which predictors we choose in the following.

Expected Relationships of Selected Predictors with Emergence of Leadership

In the following, we will provide an overview of the predictors that we use in the present research together with their respective predictions for leadership initiative and its success. We ordered them roughly thematically to each other.

Intelligence. In line with previous research (Judge et al., 2004a), we expect intelligence to be a positive predictor of leadership initiative. Intelligent persons are expected to perform well in solving complex tasks, and commonly, many leadership tasks are of such nature (see Judge et al., 2004a). Therefore, we expect intelligence to be a positive predictor of success of leadership initiative as well.

Subjective competence. We expect an individual who assesses her own competence in solving a respective task as high as more likely to show leadership initiative. Therefore, we expect subjective competence to be a predictor for leadership initiative. It is plausible that an individual, who is convinced that her ability in solving the given task is high, expresses this task-related confidence and is also successful in attracting followers. Hence, we expect subjective competence to be a positive predictor for success of leadership initiative.

Agreeableness. As agreeable individuals tend to be passive and compliant (Judge et al., 2002), we assume that agreeableness neither predicts leadership initiative, nor its success. However, we still included this trait as a predictor for leadership emergence, to assess the full range of personality measured by the big five traits, and to replicate previous research findings on this predictor.

Conscientiousness. Because conscientious individuals generally aim for achievement and in line with previous research findings (Judge et al., 2002), we expect conscientiousness to be a positive predictor for leadership initiative. Because the qualities of conscientious individuals, namely aiming for achievement and being dependable at the same time (Judge et al., 2002), should be seen as positive by followers, we also expect conscientiousness to positively predict success of leadership initiative.

Extraversion. Extraversion is considered as one of the strongest predictors for leadership emergence (Ensari et al., 2011; Judge et al., 2002). In line with these previous findings, we expect extraversion to positively predict leadership initiative. As extraversion combines characteristics that are desirable in a leader, namely being social and assertive (Judge et al., 2002), we expect extraversion to be a positive predictor for success of leadership initiative as well.

Neuroticism. Neuroticism has been reported to be negatively related to leadership emergence (Ensari et al., 2011; Judge et al., 2002). However, the evidence for the relation

between neuroticism and leadership emergence is not as clear as for the other big five traits. We rather expect neuroticism to be a negative predictor for leadership emergence. Moreover, we expect that neuroticism has a detrimental effect on success of leadership initiative so that neurotic individuals are less successful in attracting followers.

Openness to experience. As openness to experience includes the tendency to be creative (Judge et al., 2002) and was meta-analytically shown to be predictive of leadership initiative (Ensari et al., 2011; Judge et al., 2002), we expect that openness to experience is a positive predictor for leadership initiative. As openness to experience is the least understood facet of the big five and was not related to many applied criteria (Judge et al., 2002), we do not expect it to predict its success.

Self-esteem. Previous research has identified self-esteem as a significant predictor for leadership emergence (Atwater et al., 1999; Ensari et al., 2011; Paunonen et al., 2006). As making a leadership attempt involves the risk of failing (Amos & Klimoski, 2014), self-esteem might be a driving factor for showing leadership initiative. Therefore, individuals with high self-esteem might be more likely to show a leadership attempt (Acton et al., 2019). Self-esteem might be a driving factor for showing leadership initiative. Therefore, we expect self-esteem to be a positive predictor for leadership initiative. In contrast to the predicted effect of self-esteem on leadership initiative, we do not expect that an individual's internal self-view will be sufficient in successfully attracting followers.

Self-monitoring. In line with previous research, including behavioral measurements of self-monitoring (Eby et al., 2003), we predict that self-monitoring is a positive predictor for leadership initiative. Because high self-monitoring individuals are expected to adjust their behavior according to the given situation, we expect self-monitoring to positively predict success of leadership initiative, as well.

Dominance. As already shown in the literature (Ensari et al., 2011), we expect dominance to be a positive predictor of emergence of leadership. Because dominance includes qualities like being influential and persuasive (Jackson, 1974), we also expect it to be a positive predictor for success of leadership initiative.

Risk-taking tendency. Emerging as a leader can be a risk per se because making a leadership attempt carries the risk of failing (see Amos & Klimoski, 2014). However, we assume that being prone to take risks only gives a first impulse to emerge as a leader, and we do not expect effects beyond that. Therefore, we expect that risk-taking tendency positively predicts leadership initiative, but does not predict its success.

General self-efficacy. In line with previous research findings (Foti & Hauenstein, 2007; J. A. Smith & Foti, 1998), we expect that general self-efficacy is a positive predictor for leadership initiative. However, we do not assume that an individual's own expectation to perform successfully has an influence on the success of leadership initiative.

Narcissism. As narcissists are likely to tend to use situations like leaderless group situations as opportunities for self-enhancement (Brunell et al., 2008), we expect narcissism to be a positive predictor of leadership initiative. Based on the previous finding that narcissists were not successful in enforcing their opinion (Brunell et al., 2008), and because of narcissism's detrimental aspects, we expect narcissism to be a negative predictor for leadership success.

Implicit power motive. In line with research findings showing that the implicit power motive predicts the attainment of professions with high power (Winter, 1988), we expect that persons with a high implicit power motive are more likely to show leadership initiative. Because the implicit power motive includes the desire to maintain prestige and reputation, individuals with a high implicit power motive are expected to act in a way that

motivates others to follow them. Hence, we expect that implicit power motive scores positively predict both leadership initiative and its success.

Dominance motive. We expect the dominance motive to positively relate to leadership initiative. By showing leadership initiative, a person has the opportunity to influence others in positive as well as negative ways. A person might also use leadership initiative to climb the ranks in a group. Because of the negative behaviors included in the dominance motive (e.g., aggressive behaviors; Suessenbach et al., 2018), we believe that a person's dominance motive has a detrimental effect on attracting followers. Therefore, we expect that the dominance motive is a positive predictor for leadership initiative, but a negative predictor for its success.

Leadership motive. Persons may claim leadership because of their desire to lead (Suessenbach et al., 2018) and therefore, might be motivated to show leadership initiative. Additionally, the desire to lead and offering to take responsibility might make leadership attempts successful and influence others to follow that person. Because of these reasons, we expect the leadership motive to positively predict both leadership initiative and its success.

Prestige motive. We expect that the prestige motive enables persons to show leadership initiative, because a leadership attempt, especially if it turns out to be successful, can serve as an opportunity to fulfill the desire to be admired. The prestige motive is a new predictor, and we cannot derive from the concept itself, why or how it should influence the success of a leadership attempt. Therefore, we expect that the prestige motive positively predicts leadership initiative, but not its success.

Body height. We believe that body height does not have an independent effect on leadership initiative. Participants who are taller might have more self-esteem or are more confident about their own abilities and competencies and might, therefore, be more likely to show leadership initiative (Judge & Cable, 2004). Because we believe that the effect of body

height disappears when we predict leadership initiative with self-confidence and subjective competence at the same time, we expect that body height will not predict leadership initiative. However, because of evolutionary reasons, it might have been beneficial to follow a tall individual (Ronay & Carney, 2013) and therefore, we believe that body height positively predicts success of leadership initiative.

Physical fitness: Waist-to-hip-ratio. In the present research, we use participants' waist-to-hip ratio as a proxy for physical fitness in predicting emergence of leadership. Because physical fitness is related to self-esteem (Atwater et al., 1999), physically fit individuals might be more likely to show leadership initiative. However, because we believe that the effect of waist-to-hip-ratio disappears when we predict leadership initiative with self-esteem at the same time, we do not expect that waist-to-hip-ratio has an independent effect on leadership initiative. Therefore, we do not expect that the waist-to-hip ratio predicts leadership initiative. Because of evolutionary benefits associated with physical fitness, we expect that waist-to-hip ratio positively predicts success of leadership initiative.

Testosterone. As testosterone might play a role for attaining leadership positions (Van der Meij et al., 2016), we expect testosterone to be a positive predictor for leadership initiative. Because the effect of testosterone on success of leadership initiative seems rather unclear, we do not expect testosterone to be a predictor.

Perceived attractiveness. In the present research, we believe that the effect of perceived attractiveness differentially influences the two components of emergence of leadership. As outlined before, more attractive persons are supposed to be more self-confident and have a higher general self-efficacy. If we predict leadership initiative simultaneously with attractiveness and self-efficacy, we do not expect perceived attractiveness to have an independent effect. Therefore, and because we do not expect that other persons' perceptions influence whether a person shows leadership initiative per se, we

do not expect perceived attractiveness to be a predictor of leadership initiative. Applying the findings of the previously mentioned election studies, however, we predict that attractiveness is a positive predictor for success of leadership initiative.

Perceived competence. Evidence from election studies suggests that perceived competence will serve as a predictor for leadership emergence (Antonakis, Ashkanasy, & Dasborough, 2009; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). It might be plausible to assume that individuals who are perceived as competent might also make experiences that are increasing their self-esteem, for example, winning an election. If we predict leadership initiative with perceived competence and indicators of self-worth simultaneously, we do not expect perceived competence to have an independent effect on leadership initiative. As the results from election studies suggest, we expect perceived competence to predict success of leadership initiative.

Perceived trustworthiness. Although the findings on perceived trustworthiness and election success are not as clear as for perceived attractiveness and competence (Berggren et al., 2010; Todorov et al., 2005), additional findings from business contexts suggests that perceived trustworthiness is a predictor for leadership emergence (Gomulya et al., 2017; Linke et al., 2016). Individuals who are perceived as trustworthy could also gain higher views on their self-worth by experiencing that they are trusted. Therefore, they might be inclined to show leadership initiative. Because we expect this effect to disappear when simultaneously predicting leadership initiative with self-confidence, we do not expect that perceived trustworthiness has an independent effect on leadership initiative. Because it may have been proven to be evolutionarily beneficial to follow trustworthy individuals (Van Vugt & Ronay, 2014), we expect perceived trustworthiness to predict success of leadership initiative.

Additional predictors, for exploratory purposes. For exploratory purposes, we investigated the effect of additional predictors on leadership initiative and its success. These

include the explicit power motive, body weight, biacromial shoulder width, sub-facets of narcissism, measured with the narcissistic personality inventory, and a more recent measure of narcissism, the narcissistic admiration and rivalry questionnaire, including its two sub-facets admiration and rivalry.

Concluding overview of hypotheses. Table 1 provides a summary of the predicted relationships between the investigated predictors and leadership initiative and success of leadership initiative.

From the overview of identified predictors for emergence of leadership, we refrain from including the following measures. This is due to the following reasons. We exclude measures of emotional competency, intelligence, and recognition because the validity of the measures seems to be unclear (Côté et al., 2010). We do not include measures of masculinity/femininity/androgyny (gender role), because the inventories measuring these constructs are dated, and therefore, do not seem to be contemporary anymore (Kolb, 1997). Besides, we do not include race, because we conducted our study in a small city in Germany and did not expect substantial differences in the race in our sample. For the predictor gender, we investigated leadership emergence in all-male, all-female, and mixed-gender groups. Because of our research design, we only tested the effect of an individual's gender on leadership initiative and its success in mixed-gender groups.

Table 1
Summary of the predicted relationships between predictors and leadership initiative and its success

Predictor		Leadership initiative	Success of leadership initiative
Intelligence		positive	positive
Subjective competence		positive	positive
Big 5	Agreeableness	no	no
	Conscientiousness	positive	positive
	Extraversion	positive	positive
	Neuroticism	negative	negative
	Openness to experience	positive	no
Self-esteem		positive	no
Self-monitoring		positive	positive
Dominance		positive	positive
Risk-taking tendency		positive	no
General self-efficacy		positive	no
Narcissism		positive	negative
Implicit power motive		positive	positive
Dominance motive		positive	negative
Leadership motive		positive	positive
Prestige motive		positive	no
Body height		no	positive
Waist-to-hip-ratio		no	positive
Testosterone		positive	no
Perceived attractiveness		no	positive
Perceived competence		no	positive
Perceived trustworthiness		no	positive

Note. positive = positive predictor; negative = negative predictor; no = no relationship hypothesized.

Conclusions

To sum up, we identified several research goals that we want to investigate within the present research. We differentiate emergence of leadership into two components: leadership initiative and its success in attracting followers. To predict leadership initiative and its success, we selected a broad range of predictors, including predictors already known from previous research as well as newly identified predictors. Furthermore, we aim at measuring leadership initiative and its success on a behavioral level. Therefore, the group task in the present research needs to give us the opportunity that a behavioral measurement is viable, and that leadership initiative can be differentiated from success of leadership initiative. Moreover, we experimentally manipulate the gender compositions of the groups to explore whether group gender composition influences the effect of inter-individual differences on leadership emergence. For that purpose, participants either worked in mixed-gender, male, or female groups. To reach our research goal, we conducted two different studies. In this way, we were able to measure leadership initiative and its success behaviorally and very unambiguously in both studies but making the group interaction more realistic by introducing a group discussion in Study 2. We present the studies in the following.

Method

We conducted two studies to test our hypotheses. In this way, we were able to test them across varying group situations in two samples. Both studies followed the same logic, and we present their methods combined in the following. In both studies, we first collected data on the predictor variables before measuring our criterion variables during a group interaction. In each of the studies, participants took part in two sessions. During the first session, we collected data on the majority of the predictor variables. At least two days later, participants came to the laboratory for the measurement of the remaining predictor variables and our criterion variables leadership initiative and success of leadership initiative. Although

from a theoretical point of view, the "cleanest" solution would have been to measure all predictor variables during the first session, we abstained from doing so due to pragmatic concerns: For example, to maximize the predictive value of the testosterone assessment, we took the saliva samples right before we assessed our dependent variables. Participants worked in ad-hoc groups on an estimation task (Study 1) or wilderness survival task (Study 2), each with several items, and made a group decision for each item. During the group task, we assessed leadership initiative and its success on a behavioral level. Following the group task, participants were asked to answer a set of questions regarding the group setting and working as a group. All administered questionnaires and tasks were presented on a computer using the experimental software Alfred (version 0.2b5; Treffenstaedt & Wiemann, 2018), based on the programming language Python (version 2.7). All used open-source materials can be found within the ECOSOP data repository on the Open Science Framework (https://osf.io/f7wj9/). We obtained ethics approval for both studies from the Georg-Elias-Müller-Institute of Psychology's Ethics Committee (no. 145).

Study 1

Sample and design. From May to December 2016, 396 participants (96.21% students) took part in a study advertised with the title "Inter-individual differences and behavior in newly formed groups" in exchange for 35€ and an additional performance-based bonus payment of up to 15€ per person. On request, participants also received a 3D-printed 4″ figurine of themselves. People were excluded from participation if they did not speak German natively. We applied this criterion to make sure that the participants' behavior during the group interaction is not influenced by communication difficulties. Because of the hormonal assessment, we also excluded persons who were younger than 18 or older than 35 years, who suffered from an endocrine disorder, or who were currently taking antidepressants, steroids or anabolic substances. We excluded data from one group from the

analyses because its members disregarded some of the group task instructions. Due to sporadically missing values, the main analyses are based on $N = 378$ participants. We excluded testosterone values for six participants because they were missing or contaminated; values for perceived attractiveness, trustworthiness, and competence were not available for four participants because their group's photos were missing; values for waist-to-hip-ratio were not available for three participants; and values for all self-report measures except the big five and the intelligence test are missing for one participant because of an error in the computer experiment. Participants' average age was 23.79 years ($SD = 3.87$). Participants worked in ad-hoc groups of four, either in all-male, all-female or mixed-gender groups consisting of two males and two females each. The resulting sample consists of 33 male groups, 32 female groups, and 33 mixed-gender groups. We made sure that participants in a group did not know each other. Study 1 constituted a one-factorial between-subjects design with the factor group gender composition (mixed-gender vs. male vs. female). However, the stated hypotheses relate to the regression coefficients between the predictors and criterion variables.

Predictor variables. All inventories were administered so that higher scores mean more marked characteristics. Items were recoded accordingly, if necessary.

General intelligence. To assess general intelligence, we used six subtests from the Intelligence Structure Test 2000 R (Form A; Liepmann, Beauducel, Brocke, & Amthauer, 2007). The test's basic module measures verbal, numerical, and figural intelligence with three subtests. Because of time constraints, we only used two subtests for each dimension: sentence completion and verbal analogies (assessing verbal intelligence), number series and numerical signs (assessing numerical intelligence), and figure selection and matrices (assessing figural intelligence). Each subtest consists of 20 tasks. In accordance with the manual, participants received one point for each correctly solved task and could reach scores from 0 to 20 for each

subtest. Because we did not administer the complete intelligence test, we conducted a principal component analysis to obtain a general measure of intelligence (*g* factor) for each participant. In this analysis, we included all subtest scores from all participants. Only one factor had an eigenvalue over Kaiser's criterion (> 1), suggesting that there is only one strong first factor. Therefore, we extracted the first, unrotated factor and subsequently received the resulting factor score for each participant.

Subjective competence. Right before the group interaction, participants worked on three items to get familiar with the task they would work on in the group. We asked the participants to assess their competence in solving these items on a 7-point Likert scale from 1 (*not good at all*) to 7 (*very good*).

Big Five. To measure the big five traits, we used the Big Five Inventory (BFI; John, Naumann, & Soto, 2008; German version by Lang, Lüdtke, & Asendorpf, 2001). The BFI consists of 42 items and measures extraversion (8 items), agreeableness (8 items), openness to experience (10 items), conscientiousness (9 items), and neuroticism (7 items). The items were rated on 5-point Likert scales from 1 (*totally disagree*) to 5 (*totally agree*). For each personality dimension, we calculated mean values.

Narcissism. To measure narcissism, we used the Narcissistic Personality Inventory (NPI; Raskin & Hall, 1979; Raskin & Terry, 1988; German version by Schütz, Marcus, & Sellin, 2004). The NPI consists of 40 statement pairs. For each pair, the participants were asked to choose the statement they identify with the most. According to the selected statements, they received either 0 or 1 point, resulting in a total sum score between 0 and 40.

Self-monitoring. Self-monitoring was assessed using the Self-Monitoring Scale (Lennox & Wolfe, 1984; German version by Schyns & Paul, 2014). Participants were asked to indicate how often they engage in social situations described by 13 items on 6-point Likert scales from 1 (*always wrong*) to 6 (*always true*). We deviated from the manual by reversing

the original order of the answer options. In this way, answers for all questionnaires were assessed from disagreement to agreement, in order to avoid confusion and erroneous answers from the participants. Total scores (e.g., Ellis, 1988; Foti & Hauenstein, 2007) were computed, with a possible range from 13 to 78.

Self-esteem. Self-esteem was measured with the Rosenberg Self-Esteem Scale (Rosenberg, 1965; German version by Von Collani & Herzberg, 2003), consisting of 10 items. Participants answered on 4-point Likert scales from 1 (*strongly disagree*) to 4 (*strongly agree*) and we subsequently calculated mean scores.

General self-efficacy. To assess general self-efficacy, we used the General Self-Efficacy Scale (GSE; Sherer & Adams, 1983; German version by Schwarzer & Jerusalem, 1999). The scale consists of 10 items on 4-point Likert scales from 1 (*strongly disagree*) to 4 (*strongly agree*). As prescribed, we calculated a mean score.

Dominance. Dominance was measured using the dominance subscale from the Personality Research Form (PRF; Form KB; Jackson, 1967, 1974; German version by Stumpf et al., 1985). The scale consists of 16 statements. Participants had to disagree or agree with these statements, receiving either 0 or 1 point for each answer. Again, we deviated from the manual and reversed the order of the answer options to provide a consistent answer format for the participants. Dominance scores had a possible range from 0 to 16.

Risk-taking tendency. To assess risk-taking tendency, we used the Domain-specific Risk-taking Scale (DOSPERT-G; Weber, Blais, & Betz, 2002; German version by Johnson, Wilke, & Weber, 2004). For 40 different risky behaviors, participants had to indicate how likely they are to engage in each behavior on 5-point Likert scales from 1 (*very unlikely*) to 5 (*very likely*). We calculated a mean score over all items (in accordance with Foster, Shenesey, & Goff, 2009).

Dominance, prestige, and leadership motive. To assess the dominance, prestige, and leadership motive, we used the short version of the dominance, prestige, and leadership motive scales (Suessenbach et al., 2018). Together with the original first author of the scales, we translated the scale to German. The short scale consists of 18 items, six for each facet. The items consist either of statements or goals. Participants indicated their agreement with statements on 6-point Likert scales, ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). For goal items, participants indicated how important the described goals are for them, as well on 6-point Likert scales, with a possible range from 1 (*not important at all*) to 6 (*very important*). In accordance with the manual, the scores were recoded to range from 0 to 5. We calculated sum scores for each facet with a possible range from 0 to 30.

Implicit power motive. We assessed the implicit power motive using the Picture Story Exercise (PSE). In the PSE, several ambiguous pictures are presented, and participants are asked to write imaginative stories about these pictures. Subsequently, the written stories are scored according to the amount of motive-related imagery, defined as the attribution of an action, wish, concern, or internal state to others (Winter, 1994). The amount of imagery related to each motive serves as an approximation for the strength of the implicit motive (Schönbrodt et al., 2018). In general, the PSE can be used to assess the implicit power, affiliation, and achievement motive. Because we focused on the implicit power motive, we followed a recommendation to use pictures aiming at the targeted motive (Schönbrodt et al., 2018) and chose pictures having a high pull on power imagery. We used two pictures from the standard set for administering the PSE (ship captain and women in the laboratory; Schultheiss & Pang, 2007) as well as three new pictures (applause, beachcombers and Kennedy Nixon; Schönbrodt et al., 2018). We pretested the new pictures ($N = 20$), ensuring that they have a strong pull on power imagery (see Schönbrodt et al., 2018, for comparable pulls). In the main study, participants worked on the five pictures following standard

instructions (Schultheiss & Pang [2007]; German translation adapted to administration via computer). Each picture was shown for 10 seconds on the computer screen. Subsequently, participants were directed automatically to the next screen where they were instructed to write their story. On this screen, guiding questions (Schultheiss & Pang, 2007) and a small version of the previously shown picture were presented. Participants were instructed to write a story within four minutes. The pictures were presented in a randomized order, and stories were coded according to the standard procedure suggested by Winter's (1994) *Manual for Scoring Motive Imagery in Running Text*. Following the manual, need for power (*n* Power) is scored when a story character has impact, control or influence on others or the world at large, for example by acting strongly and forcefully, controlling and regulating, influencing or persuading, giving unsolicited help or advice, or eliciting strong emotions in someone (Winter, 1994). According to this manual, the implicit motives are scored sentence-wise. Therefore, each sentence can be scored for the presence of the implicit power, achievement, or affiliation motive with 0 (the respective motive is not present) or 1 (the respective motive is present). We deviated from the manual in two ways. First, instead of scoring one motive several times in one sentence, we only allowed scoring each motive once within one sentence. Second, we abandoned the "2nd-sentence-rule", which states that the same motive cannot be scored in two subsequent sentences (see Schönbrodt et al., 2018). For each participant, we counted the number of scores per motive. Scores range from 0 to 34. Stories were coded mainly by one trained coder at Ludwig-Maximilians-University Munich who was blind to our hypotheses. The coder reached high interrater reliability with expert coded material, $ICC > .85$. Although the scores for all three motives were coded, we will only report scores for the implicit power motive. The other motives were not in the scope of our hypotheses. Motive scores and word count for each person were significantly correlated, $r = .65$, $p < .001$, and we controlled for word count by residualizing motive scores for word count

and extracting residual scores per person. We used these scores in all following analyses (in accordance with Schultheiss & Pang, 2007).

Testosterone. Testosterone was measured in picograms per milliliter using saliva samples. To minimize contamination, participants were asked to avoid exercising, consuming alcohol, or any kind of drugs the day of the second session, refrain from caffeine three hours before the appointment, and avoid eating or drinking (except for water), brushing their teeth or smoking cigarettes one hour before the appointment (Kordsmeyer, Hunt, Puts, Ostner, & Penke, 2018). Samples were provided between noon and 6 pm to control for circadian fluctuations (Schultheiss & Stanton, 2009). Before providing a saliva sample, participants completed a hormone screening questionnaire (Schultheiss & Stanton, 2009). This questionnaire can be used to assess potentially biasing influences on the samples, for example, whether participants take medicine. Participants were asked to provide at least 1 ml saliva via unstimulated passive drool through a straw (Schultheiss et al., 2012). Samples were examined for any visible signs of blood traces and taken again if necessary. Samples were stored immediately in an ultra-low temperature freezer at -80°C. After finishing the data collection, samples were packed on dry ice and subsequently shipped to the lab of Clemens Kirschbaum at the Technical University of Dresden. They were analyzed using high sensitivity immunoassay with chemiluminescence detection (CLIA, IBL-International, Hamburg, Germany). Measurement precision of salivary hormones is commonly evaluated as the coefficient of variation (CV). The CV is defined as the mean of replicate measurements of a sample, divided by the standard deviation of the measurements and multiplied by 100. Furthermore, it is commonly differentiated between the intra- and inter-assay coefficient. The intra-assay coefficient is defined as the mean CV of all given samples in an assay. The inter-assay CV is calculated from the between-assay mean and *SD* from a control sample. Generally, intra- and inter-assay CVs below 10% are considered desirable (Schultheiss &

Stanton, 2009). In this sample, the intra-assay CV is below 5%. The inter-assay CV is determined with control samples provided together with the analysis kit by IBL-International and is below 7%. Data from two participants were missing, one sample was contaminated and could not be analyzed, and another three participants claimed to take hormonal medication, so we excluded the hormone data from these cases. Five participants reported to have endocrinologic disorders, but because it remained unclear whether their testosterone levels were biased, we decided to keep the hormone data in these cases in the analyses. We winsorized outliers to 3 *SDs* to their gender means, that is, assigning this value to these outlying cases ($n = 7$; five males, two females; Mehta, Welker, Zilioli, & Carré, 2015). The resulting testosterone values were positively skewed and, therefore, not normally distributed (Shapiro-Wilk test $W < 0.95$, $p < .001$). The testosterone values were log10-transformed to normalize the data for subsequent analyses (Mehta et al., 2015).

Physical features. We created 3D body scans using a Vitus Smart XXL 3D body scanner, running Anthroscan software (both Human Solutions GmbH, Kaiserslautern). We used the body scanner in collaboration with the lab of Lars Penke at the Georg-August University of Goettingen. Each participant was scanned three times. For the scans, participants wore tight and standardized underwear, removed jewelry, and tied up their hair if necessary. Participants were instructed to strike a standard pose according to the manufacturer's recommendations (standing upright with legs hip-wide apart, head in accordance with the Frankfort Horizontal to make sure the participant's head was parallel to the floor, arms angled slightly, breathing normally). All measures were averaged across the three scans.

Waist-to-hip-ratio. Waist-to-hip-ratio was calculated by dividing each participant's mean waist girth by their mean hip girth (both in cm). Reliabilities for the waist and hip girth were high, both *ICCs* = .99.

Body height. We measured participants' body height once using a stadiometer (in m). Participants were barefoot.

Perceived attractiveness, competence, and trustworthiness. After the data collection was finished, we ran a rating study to obtain average perceived attractiveness, competence, and trustworthiness ratings for each participant. External student raters ($N = 94$) rated photos of participants' faces that were presented showing a neutral expression. Photos were presented groupwise, and each rater saw a subset of the full sample with an equal amount of male, female, and mixed-gender groups. We decided to split the sample into three subsets to give the raters a reasonable number of persons to rate, making sure that each participant was rated by at least $n = 30$ raters. Raters were asked to score each participant's attractiveness, competence, and trustworthiness relative to the other group members on 7-point Likert scales from 1 (*not at all*) to 7 (*a lot*)[2]. Ratings for each participant and dimension were averaged (see Klimpki, 2017, for a detailed description of the rating study).

Accuracy of proposed answers. Because it would represent a highly reasonable strategy to follow accurate answers, the accuracy of proposed answers was expected to have a critical impact in predicting whether a proposed answer was accepted as the groups' decision. For that reason, we calculated the mean accuracy of the suggested answers for each participant. Values ranged from 0 (proposed answers were never correct) to 1 (proposed answers were always correct) and were only used for predicting success of leadership initiative.

---

[2] We aimed at measuring the ratings relative to the other group members, because an individual's perceived attractiveness might have an influence on our criterion variables depending on the attractiveness of the other group members. For example, an average attractive person might be perceived as the most attractive person in a group with relatively unattractive group members but might be perceived as the least attractive person in a group with relatively attractive group members.

Follow-up questions.

Suspicion check. We asked the participants to indicate the aim of the study in an open answer format.

Perceived leadership initiative. We asked the participants to rate each group member, including themselves with regard to the degree they perceived that this person proposed answers during the group tasks. Ratings were made on a 7-point Likert scale from 1 (*not at all*) to 7 (*a lot*).

General leadership impression. The general leadership impression scale (GLI) is a classic tool to measure leadership perceptions (Cronshaw & Lord, 1987). We asked the participants to assess their group members and themselves on five items according to their appearance as a leader on 5-point Likert scales from 1 (*not at all*) to 5 (*a lot*). We calculated sum scores for each participant, with a possible range from 20 to 100.

Leader nomination measure. Another frequently used measure of leadership perceptions is the leader nomination measure (Eby et al., 2003; Zaccaro et al., 1991). We asked the participants to imagine they would work with their group a second time. Then they should rank their group members, including themselves, with regard to the extent that they would like them to be a leader for the group. Ranks were assigned in decreasing order (i.e., the group member they preferred the most as a group leader was ranked as number 1, and the group member they preferred the least to be the group's leader was ranked as number 4). We reversed the item for data analyses so that higher values meant a person is preferred more. We calculated a mean ranking score for each participant by averaging the rankings each participant received from her group members and herself.

Motivation during the group task. We asked the participants several questions about their motivation during the work in the group. Six questions were answered on 7-point Likert

scales from 1 (*not at all*) to 7 (*a lot*). We subsequently averaged all six ratings into a mean score.

Group identification. To measure an individual's identification with the group, we used the group identification scale (Doosje, Ellemers, & Spears, 1995). We asked the participants to indicate their agreement to four statements about the identification with the group on 7-point Likert scales from 1 (*strongly disagree*) to 7 (*strongly agree*). We calculated the mean scores for each participant.

Comments. Participants had the opportunity to comment on the study and to give feedback about how we conducted the study.

Additional predictor variables, for exploratory purposes.

Narcissism (sub-facets of the Narcissistic Personality Inventory). Although it is common practice to use the total score of the Narcissistic Personality Inventory as a measure of narcissism, current research suggests that narcissism is a construct consisting of adaptive and maladaptive facets. Because a total score might blend these together (Ackerman et al., 2011; Back et al., 2013), we aimed at exploring the relation between sub-facets of narcissism and emergence of leadership following a current factor solution of the Narcissistic Personality Inventory (Ackerman et al., 2011). We investigated the following sub-facets.

Leadership/Authority. This facet contains eleven statement pairs of the total inventory. According to the selected statements, participants received either 0 or 1 point, resulting in a total score between 0 and 11.

Grandiose Exhibitionism. This facet contains ten items of the total inventory. According to the selected statements, participants received either 0 or 1 point, resulting in a total score between 0 and 10.

Entitlement/Exploitativeness. This facet contains four items of the total inventory. According to the selected statements, participants received either 0 or 1 point, resulting in a total score between 0 and 4.

Narcissism (Narcissistic Admiration and Rivalry Questionnaire). In addition to the Narcissistic Personality Inventory, the traditionally used measure for narcissism, we used the more current Narcissistic Admiration and Rivalry Questionnaire (NARQ; Back et al., 2013). Participants indicated their agreement to 18 statements on 6-point Likert scales, ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). For an overall measurement of narcissism, we computed a mean score. Additionally, we used the inventory's sub-facets, admiration and rivalry, consisting of 9 items each. We calculated mean scores for these sub-facets.

Explicit power motive. To assess the explicit power motive, the short version of the power scale of the Unified Motive Scale (UMS; Schönbrodt & Gerstenberg, 2012) was used. The scale contains six items, which either describe statements or goals. Participants indicated their agreement with statements on 6-point Likert scales, ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). For goal items, participants indicated how important these goals are for them, as well on a 6-point Likert scale, with a possible range from 1 (*not important at all*) to 6 (*very important*). In accordance with the manual, the scores were recoded to range from 0 to 5. We computed a mean score.

Body weight. Body weight (in kg) was assessed using a scale integrated into the 3D-body scanner.

Biacromial shoulder width. Biacromial shoulder width (in cm) was measured by an undergraduate assistant using the participants' 3D-body scans. The reliability of the measurements was high, *ICC* = .99. We averaged the biacromial shoulder width over the three scans.

Criterion variables.

Leadership initiative. During the group phase, participants worked together on several estimation items. In order to solve these items, one group member had to propose an answer and give an explanation in a first step. In a second step, the other group members voted whether that answer was accepted as the group's decision. Leadership initiative was operationalized as the number of times a group member proposed an answer and explained why it might be correct in order to convince her group members to vote for it. Values for leadership initiative could possibly range from 0 to 15.

Success of leadership initiative. In order to determine success of leadership initiative, we first counted the number of group members following (by voting for) a participant's proposed answer in each trial. According to the rules of the tasks, at least two group members (in addition to the one who had proposed the solution) had to follow the proposed answer to be accepted as the group's decision. We measured success of leadership initiative for each trial with values of either 0 (i.e., the proposed answer was rejected as the group's decision) or 1 (i.e., the proposed answer was accepted as the group's decision). To maximize validity, we decided to calculate a robust measure of success of leadership initiative. Participants who showed leadership initiative only once either might have felt an obligation to propose an answer at least once regardless of its correctness or thought they knew the answer for only this one item. In these cases, our measure of success of leadership would base on only one ambiguous observation. With these considerations in mind, we decided to calculate a more robust measure. In addition to excluding participants that never showed leadership initiative, we excluded all cases in which participants only showed leadership initiative once (2.40% of the trials). We used the resulting variable for all main analyses. For correlation analyses with the other variables used in this study, we calculated each participant's success ratio. For that purpose, we divided their number of successes by the number of times they showed

leadership initiative, with a possible range from 0 (a participant never had success) to 1 (all proposed answers were accepted by the group).

Additional criterion variable, for exploratory purposes.

Overall leadership. In practice, leaders are persons who show leadership initiative and are successful with their attempts. That is why we aggregated leadership initiative and its success to obtain an overall measure of leadership. We operationalized overall leadership as whether a participant showed leadership initiative, and the group also accepted the proposed answer as its decision. Overall leadership is calculated on trial level with values of either 0 (a participant did not show leadership initiative in this trial or a participant showed leadership initiative, but the proposed answer was rejected) or 1 (a participant showed leadership initiative, and her answer was accepted). We used that variable for the main analyses. To validate overall leadership with the general leadership impression scale (GLI), we computed total scores for each participant.

Materials. For us to be able to investigate the effect of our chosen predictors on leadership initiative and its success, the group task had to fulfill several general criteria. First, we aimed at measuring leadership initiative and its success behaviorally and unambiguously over several trials. Second, we aimed at testing as many trials as possible. Third, participants should not possess professional expertise in the task. Professional expertise might overrule the effect of predictor variables and might motivate a person to show leadership initiative. Additionally, it is highly plausible that participants follow an expert independent of her individual characteristics. Fourth, we aimed at using tasks with a correct answer, enabling us to pay the participants according to their group's performance. In tasks without an objectively correct answer as, for example, choosing the most beautiful painting among several works of art, we cannot incentivize participants according to their objective performance. Fifth, participants should be able to explain why their proposed answer is reasonable, but without

being able to prove the correct answer. In tasks with maximum demonstrability, the group members might propose an answer because they already know it is correct. Additionally, we would expect the other group members to agree or disagree with a proposed answer solely because they already have seen whether it is correct or not. As a specific criterion for this study, we wanted to use a relatively difficult task. We were interested in predicting leadership initiative and its success in situations with high insecurity. In this way, we aimed at ruling out any confounding effects that might motivate persons to propose or agree to an answer.

Given these requirements, we decided to use an estimation task with items from different topics with four answer options. For example, participants were asked to estimate the height of a person and had four possible answer options. To adjust the material to our needs, a picture of the object that participants should estimate accompanied each item (see Figure 1 for an example item). We developed several items and tested them in a series of pretests. Initially, we developed 52 items. To ensure that we have items from different categories (e.g., estimating the distance between two cities or estimating a person's body weight) in the group phase, we developed, if possible, two items from each estimation category (e.g., two different items asked to estimate a person's height). In a first, individual pretest ($N = 47$), we asked participants to provide a numerical estimation for each item. For example, in the case of the item shown in Figure 1, we asked the participants to estimate the depicted person's height in cm. In this way, we received a range of values that participants consider plausible. This information was used to develop the answer options for each item. To make sure we select difficult items, we compared participants' estimation errors for items from the same category. For that purpose, we calculated the mean absolute percentage error (MAPE) for each item. High MAPE values indicated a larger estimation error, and, therefore, greater difficulty. For items belonging to the same estimation category, we eliminated the one with the smaller MAPE for further testing. Additionally, we eliminated items whose answer

distribution indicated that they violate our task criteria. For example, the distributions of provided estimates for some items were very narrow and scattered closely around the correct answer, making it extremely difficult to develop a set of four reasonable answer options. According to these rules, 28 items qualified for further testing.

Figure 1. Example item of the group task in Study 1. From: http://www.estimation180.com/; License: https://creativecommons.org/licenses/by-nc-sa/3.0/; The original work has been modified.

We developed possible answer options for the remaining pool of items. We aimed at choosing items for which participants' answers in the pretest were, approximately, distributed equally over at least three of the four options we provided. We selected items achieving that benchmark after pretesting different sets of answer options in two pretests ($N = 20$ and $N = 19$). During both pretests, participants worked individually and were instructed to choose the correct answer for each item (see the online supplementary material for selection rates). We eliminated seven items in which the answer distributions were not roughly distributed over at least three of the four answer options in at least one of the pretests.

In a fourth pretest, we wanted to make sure that participants were not able to recognize an item's correct answer when it was proposed (resembling a eureka effect). In this pretest, participants worked individually on each of the items as follows. To represent the situation in a group, we showed the participants the available answer options for each item, but only one of these four original answer options was proposed. Participants then had to indicate whether they believed that answer to be correct or not. To test every available answer option for each item, we created four separate conditions and proposed a different answer option in each. Participants were randomly allocated to the conditions (n = 10 per condition). It is important to note that we did not expect that the majority of the participants would dismiss the correct answer for every single item. The aim of this pretest was to make sure that participants did not recognize the correct answer in the majority of the cases. However, we expected that the majority of the participants would agree with the correct answer for some items. Even by chance, the participants would agree to the correct answer in a fourth of the cases. At the same time, we wanted to make sure that our participants had the possibility to solve the tasks correctly. To account for this tradeoff, we chose 15 items. Our results showed that for five out of these fifteen items, the majority of the participants agreed to the correct answer when it was proposed (see the online supplementary material for answer proportions of correct and incorrect answer options). In sum, we developed 15 items meeting our criteria after this pretest series.

Procedure. In this study, both experimental sessions took place at the lab. For the first session, we invited up to twelve participants. When arriving at the lab, participants were welcomed by two experimenters, informed about the procedure, and signed an informed consent. Then, participants completed the self-report personality inventories individually, starting with the Big Five Inventory to warm up and followed by the intelligence test. Subsequently, participants worked on all remaining self-report personality inventories. We

administered all inventories using the same Likert scales together.[3] Participants were instructed to raise their hand once they finished a part of the inventories or the intelligence test, so we could take the 3D body scan and the photos without interrupting them while currently working on one inventory. After the participants completed all measures, they received an appointment for the second session that took place at least two days later.

During the second session, we invited groups of four participants to the lab. Upon arrival, participants were welcomed and instructed about the procedure by the experimenter. Participants first completed the Picture Story Exercise (for approximately 30 minutes) and the hormone screening questionnaire (Schultheiss & Stanton, 2009), before providing a saliva sample. Subsequently, participants were informed that they were about to solve estimation items in a group. To get familiar with the task, participants solved three estimation items individually and assessed their own competence in solving these tasks. During the following task, participants sat in a semi-circle and cooperatively worked on 15 estimation items. We presented the items in a randomized order on a TV located approximately two meters in front of the group. Participants received instructions to work on the estimation items as follows. If a group member wanted to solve an item, she was supposed to raise her hand, go to the TV, propose an answer, and explain why she thinks her proposition is correct. As already described, this constituted our measure of leadership initiative. After the explanation, the remaining group members were supposed to indicate whether they agree with the proposed answer by showing either a green card to signal agreement (i.e., following) or a red card to signal disagreement (i.e., not following). At least two group members had to agree to the

[3] The participants worked on the inventories for the respective predictor variables in the following order: self-monitoring, narcissistic admiration and rivalry together with the explicit power motive, and the dominance, prestige, and leadership motive, followed by risk-taking tendency, narcissistic personality inventory, dominance, and finally self-esteem together with general self-efficacy.

proposed answer to accept it as the group's decision (i.e., success of leadership initiative). It is important to note that neither had participants the opportunity to choose a different answer option than the one proposed, nor did they get feedback regarding the accuracy of the proposed answers. We calculated the reached bonus in each trial, summed it over all trials, and applied it to the whole group. To motivate the participants to propose a correct answer and to judge correctly whether a proposed answer is correct, we introduced a twofold bonus scheme: If a group member proposed the correct answer and the group accepted it, the group received a bonus of 1€ for that trial. Additionally, if a group member proposed an incorrect answer and the group dismissed it, the group got a bonus of 0.25€ for that trial. In any other case, the group received no bonus. After completing the group task, participants answered questions about the work in the group, were debriefed, thanked, and received their compensation.

Exclusion of trials. If participants did not comply with the instructions for the group task in a given trial, we excluded data from this trial from the subsequent analyses. Following this rule, we removed data from altogether 12 trials because, in these trials, participants proposed an answer they considered incorrect and explicitly asked the other group members to disagree. In these cases, the participants did not propose an answer they considered correct, thereby contradicting both our operationalization of leadership initiative and our experimental instruction. Additionally, in such cases, the other group members could either indicate disagreement because it was requested or because they truly believed the answer is incorrect. Therefore, we removed data from these trials for subsequent data analyses.

Study 2

We conducted a second study to pursue two main goals. First, we aimed at making the work in the group more realistic within our paradigm. Although the highly pre-structured task and relatively artificial group interaction in Study 1 offer the advantage of having a high

amount of control over the group interaction and, thereby, measuring leadership initiative and its success unambiguously, it neglects that working groups usually discuss propositions while solving tasks. For that reason, we gave the participants the opportunity to discuss the proposed answers. Second, we aimed at replicating the results of the first study with another group task, namely tasks on how to survive best in the wilderness. We chose this task because we believed that the participants could engage more easily in a discussion about the content. The structure of the wilderness survival task is comparable to the estimation tasks used in Study 1. As in Study 1, we conducted two experimental sessions. For pragmatic reasons, the first session took place online. We conducted the second experimental session in the laboratory, including the group phase.

Sample and design. From August 2017 to March 2018, 404 participants (98.27% of whom were students) took part in the second study. Participants received 35€ and an additional performance-based bonus payment of up to 12€ per person. In contrast to Study 1, participants optionally received a file with a 3D-image of their body instead of a 3D-printed figurine. In the second of two experimental sessions, participants worked in ad-hoc groups of four. We excluded data from two groups from the data analyses because these groups did not finish the group task. Additionally, we excluded data from another four groups because one of their group members did not speak German natively.[4] The resulting sample consists of 33 male groups, 31 female groups, and 31 mixed-gender groups. Due to sporadically missing values, the main analyses are based on $N = 376$ individuals (we excluded testosterone values for four participants because they were contaminated). Additionally, there were missing values for two participants for the leader nomination measure, the measure of motivation during group task, and the group identification measure, because of a technical error in the

[4] Because, prior to the second session, communication with the participants only took place via email, we were not able to pre-assess the participants' German skills reliably in every case.

computer experiment. These cases did not affect the number of observations for the main analyses. Participants' average age was 22.83 years (*SD* = 3.35). Study 2 used the same design as Study 1, that is, a one-factorial between-subjects design with the factor group gender composition (mixed-gender vs. male vs. female). Again, the hypotheses relate to the regression coefficients between the predictors and dependent variables.

Predictor variables. We collected data for the same measures as in Study 1.[5] As in Study 1, we point out some sample specific information on a few measures that we report in the following.

General intelligence. To obtain a general measure of intelligence, we again conducted a principal component analysis. Within this analysis, two factors had an eigenvalue over Kaiser's criterion (> 1). To follow a consistent procedure, we decided to only extract the factor with the highest eigenvalue as a measure for participants' general intelligence.

Implicit power motive. The scores for *n* power range from 0 to 24 and were significantly correlated with word count, $r = .65$, $p < .001$, and we controlled for word count by residualizing motive scores for word count and extracting residual scores per person.

Testosterone. For the testosterone assessment, the intra-assay, as well as inter-assay coefficients, are below 7%. Data from four participants were excluded because they were taking hormonal medication, and one participant claimed to have an endocrine disorder, but again, we decided to include this case in the analyses. Outliers were winsorized to 3 *SDs* to their gender means (*n* = 3; two males, one female). Testosterone values were positively

[5] For exploratory purposes, we assessed the long version of the explicit power motive and dominance, leadership, and prestige motive. All included measures were assessed with ten items compared to six items in the short version. Like in Study 1, we based the analyses for both studies on the short version of these scales.

skewed and therefore not normally distributed (Shapiro-Wilk test $W < 0.94$, $p < .001$). To normalize the data for subsequent analysis, they were log10-transformed (Mehta et al., 2015).

Waist-to-hip-ratio. For the measurements of both waist and hip girth, we had high reliabilities, both *ICCs* = .99.

Perceived attractiveness, trustworthiness, and competence. For the ratings of participants' perceived attractiveness, trustworthiness, and competence, we ran a rating study with $N = 90$ external student raters. Again, we divided the sample into three subsets for the raters. In this study, we matched the raters' and participants' gender, that is, female raters rated female groups, male raters rated male groups, and an equal amount of female and male raters rated mixed-gender groups to approximate the perceptions in the interacting groups better. Each participant was rated by at least $n = 29$ raters (see Lovric, 2018, for a detailed description of the rating study).

Additional predictor variables, for exploratory purposes.

Biacromial shoulder width. For the measurement of the biacromial shoulder width, we also had high reliability, *ICC* = .93.

Criterion variables.

Leadership initiative. During the group phase, participants worked together on wilderness survival tasks. They worked on them following the same procedure as in Study 1. Again, leadership initiative was operationalized as the number of times a group member proposed an answer and explained why it might be correct in order to convince her group members to vote for it. Because there were only 12 wilderness survival tasks, values for leadership initiative could range from 0 to 12.

Success of leadership initiative. As in Study 1, we measured success of leadership initiative as whether a participant's proposed answer was accepted by the group in a given trial. Success of leadership initiative was measured for each trial with values of either 0

(i.e., the proposed answer was rejected as the group's decision) or 1 (i.e., the proposed answer was accepted as the group's decision). In order to maximize the validity of this measure, we calculated a robust measure for success of leadership initiative by excluding trials from participants who showed leadership initiative only once. Therefore, we removed 4.65% of the trials.

Materials. We searched for a task which meets the same general criteria as the task described in Study 1. A task on surviving in the wilderness was suitable to our needs, and we used an already existing item set. Participants worked on 12 wilderness survival items with three answer options each (Pfeiffer & Jones, 1976). We translated the items to German. The correct answers came from the comprehensive course on woodland survival taught by the Interpretive Service, Monroe County (New York) Parks Department. For example, participants worked on the following item:

> You must ford a river that has a strong current, large rocks, and some white water. After carefully selecting your crossing spot, you should:
>
> a) leave your boots and pack on.
>
> b) take your boots and pack off.
>
> c) take off your pack, but leave your boots on.

Because the items and answer options for this task already existed, we only pretested how difficult participants perceive the items (similar to the second and third pretest in Study 1) and participants' ability to recognize an items' correct answer if it was proposed (similar to the fourth pretest in Study 1). We included all 12 items in all our pretests. We determined item difficulty with an individual pretest ($N = 29$) that had the same structure as pretests 2 and 3 in Study 1. Results revealed that the item set included roughly the same amount of easy, intermediate, and difficult items (see the online supplementary material for accuracy rates). On average, the items were roughly perceived as difficult as the items in Study 1. To

test whether participants recognized correct answers when they were proposed, we performed a pretest in which participants worked individually on the items with only one proposed answer option. This pretest had the same structure as the fourth pretest in Study 1. For each item, the participants had to decide whether the proposed answer is correct or not. Since the items on wilderness survival have only three answer options, there were only three conditions in this pretest ($N = 36$; n per condition = 11, 14, 11). For four out of twelve items, the majority of the participants recognized the correct answer (see the online supplementary material for selection rates of correct and incorrect answer options). Again, we did not expect that our participants always dismissed the correct answer and considered this result as satisfactory for our purposes. Therefore, we used all 12 pretested items for the group task.

Procedure. In contrast to Study 1, the first experimental session, where participants completed all self-report questionnaires, took place online. Participants received a link leading to the online questionnaire from us. Within this online session, participants were informed first about the study and agreed to an informed consent. The online questionnaire included the Big Five Inventory, the Picture Story Exercise, and subsequently, all remaining self-report questionnaires in the same order as in Study 1. After they completed all measures, they contacted the experimenters to receive an appointment for the second session. Two or more days later, participants came in groups of four to the laboratory for the group session. First, the experimenter welcomed the participants and informed them about the procedure, and let the participants sign the informed consent they agreed to online. In this session, participants completed an intelligence test, and we took anthropometric measures, photos, and saliva samples. Subsequently, participants were informed that they were about to solve items on wilderness survival in a group. We asked the participants to individually solve three wilderness survival items resembling the group tasks' items (developed and pretested by the first author). Subsequently, we asked the participants to assess their own competence in

solving these tasks. During the following group task, participants worked on 12 wilderness survival items. Participants sat in a semi-circle around a table where an iPad was placed upright, presenting the wilderness survival items in randomized order. We asked the participants to work on the items as follows. If a group member wanted to solve an item, she was supposed to raise her hand, reach out to the iPad to log in her proposed answer and explain why she thinks the proposed answer might be correct (i.e., showing leadership initiative). After the explanation, all group members had the opportunity to engage in a discussion about whether the proposed answer is the correct solution to the given task. We instructed the participants to avoid stating their own preference during the discussion. In this way, we wanted to prevent the participants from reaching consensus before they were supposed to indicate their decision individually. Following the discussion, every group member except the one who proposed the answer was supposed to raise a green (i.e., following) or red card (i.e., not following) to show agreement or disagreement with the proposed answer (i.e., success of leadership initiative). We asked the group member who proposed an answer to type into the iPad how many group members agreed with the proposed answer. Then, participants were shown the next item. After completing the last item, participants answered questions about working in the group, were debriefed, thanked, and received their compensation.

Test Power Analysis

Given our samples, we were interested in how likely it was to observe significant effects for leadership initiative in our studies. Because we are interested in the effects of our predictor variables over both studies, we aimed at performing an integrated analysis in the following. Using the statistic software G*Power (version 3.1.9.3; Faul, Erdfelder, Buchner, & Lang, 2009), we performed all power calculations based on a joint sample size of $N = 754$ for a linear regression model with 23 predictors. Assuming a medium effect size of $f^2 = .15$

(Cohen, 1988) and a significance level of α = .001, we reach a test power of .99. Consequently, we have a high test power assuming a medium effect size independent of the applied significance level. Assuming a small effect size of f^2 = .02, however, our test power is reduced as follows. Under a significance level α = .001, we reach a test power of .12, under a significance level α = .01, we reach a test power of .32, and under a significance level of α = .05, we reach a test power of .57. Hence, the combined analysis of our two studies can detect medium effects (and, of course, all effects that are larger than that) very confidently, while not being particularly well-suited for the detection of small effects.

Results

We performed all calculations using the statistic software R, version 3.5.1 (R Core Team, 2018). Additionally, we used the following packages in the analysis: psych 1.8.4 (Revelle, 2018), dplyr 0.7.7 (Wickham, François, Henry, & Müller, 2018), survival 2.43.1 (Therneau, 2015), and ggplot2 3.1.0 (Wickham, 2016). Because we are most interested in results that we can generalize over both studies, we focus on presenting integrative analyses for the results, including data (i.e., observations) from both studies. The results for each separate study can be found in the online supplementary material. If we find effects moderated by the study they originate from, we added the results from the study level as well.

For the main analyses, that is, performing hypothesis tests for the criterion variables leadership initiative, success of leadership initiative, and explorative analyses of overall leadership, we applied a model to the data containing all 23 predictor variables. Within this model, we included the main effects of the predictors. In the following, we refer to this model as the "predictor model". Because we focused on performing integrative analyses, we accounted for a possible moderation of the predictors' effects by the study they originate

from. Therefore, we added interaction effects between the factor study and each predictor to the predictor model, unless stated otherwise.[6]

Descriptive statistics and reliability measures (if applicable) of the measures used in Study 1 and 2 can be found in Table 2. Table 3 shows averaged correlations for the variables used in Study 1 and 2. All predictors were centered on their group means and z-standardized for the reported regression analyses.

[6] For the main analyses, we performed robustness checks. In Study 1, one participant claimed to already know the answer in a trial. We performed robustness checks excluding that trial, yielding comparable results. Additionally, some of the external raters responsible for the ratings of participants' perceived attractiveness, competence, and trustworthiness, were older than 35 years and therefore older than the participants they rated, which might have a biasing influence on their ratings (n = 4 Study 1; n = 6 in Study 2). We excluded their ratings for robustness checks, again yielding comparable results.

Table 2

Means, standard deviations, reliability coefficients for Study 1 and Study 2

Variable	Study 1			Study 2		
	M	*SD*	*α*	*M*	*SD*	*α*
1. Leadership initiative	3.72	2.12	-	3.00	1.76	-
2. Success ratio	0.70	0.24	-	0.83	0.21	-
3. GLI	54.34	13.86	.93	57.59	12.36	.92
4. Leader nomination	2.49	0.11	-	2.50	0.06	-
5. *g* factor	0.00	1.00	-	0.00	1.00	-
6. Subjective competence	3.56	1.30	-	3.77	1.24	-
7. Agreeableness	3.55	0.59	.73	3.61	0.59	.89
8. Conscientiousness	3.43	0.62	.79	3.48	0.60	.75
9. Extraversion	3.56	0.70	.86	3.46	0.78	.77
10. Neuroticism	2.87	0.76	.82	2.83	0.74	.81
11. Openness	3.72	0.62	.82	3.68	0.62	.82
12. Self-esteem	3.17	0.57	.88	3.12	0.57	.88
13. Self-monitoring	50.06	4.82	.44[a]	49.72	4.94	.41[a]
14. Dominance	9.39	3.69	.80	8.89	3.75	.80
15. Risk-taking tendency	2.80	0.48	.87	2.72	0.47	.86
16. General self-efficacy	2.94	0.43	.84	2.91	0.45	.86
17. Narcissism	14.40	6.32	.82	13.95	6.70	.85
18. *n* power	0.00	3.69	-	0.00	3.15	-
19. Dominance motive	10.43	5.42	.81	9.58	5.07	.80
20. Leadership motive	16.01	5.78	.90	15.54	5.78	.89
21. Prestige motive	17.95	4.50	.72	17.55	4.50	.70
22. Body height	1.75	0.10	-	1.76	0.09	-
23. Waist-to-hip-ratio	0.79	0.07	-	0.78	0.07	-
24. Testosterone	1.48	0.51	-	1.47	0.51	-
25. Perceived attractiveness	3.25	0.97	.96[b]	3.26	0.85	.95[b]
26. Perceived competence	4.28	0.55	.87[b]	4.35	0.51	.77[b]
27. Perceived trustworthiness	4.04	0.59	.87[b]	4.15	0.60	.82[b]

Note. GLI = General Leadership Impression. Because of missing values, sample sizes ranged from N = 386 to N = 392 in Study 1 and from N = 376 to N = 380 in Study 2. For success ratio, the sample size is N = 332 in Study 1 and N = 299 in Study 2. [a]Reliability for the total score of self-monitoring appeared low, and there are relatively low reliabilities reported in published previous research (e.g., α = .67; Zaccaro et al., 1991). [b]Reliabilities for perceived attractiveness, competence, and trustworthiness were averaged because raters rated one of three sets of participants. In Study 1, reliabilities for perceived attractiveness were α = .96 for all sets; but ranged for perceived competence from α = .86 - .89 and ranged for trustworthiness from α = .86 - .88. In Study 2, reliabilities for perceived attractiveness were α = .95, and for perceived competence α = .77, but ranged for trustworthiness from α = .81 to .82.

Table 3

Averaged correlations for variables used in Study 1 and 2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1. Leadership initiative																										
2. Success ratio	.16**/																									
3. GLI	.72**/**	.24**/**																								
4. Leader nomination	.05	.03	.05																							
5. *g* factor	.14**/	.13*/	.18**/	.01																						
6. Subjective competence	.23**/**	.02	.25**/**	-.03	.08*/																					
7. Agreeableness	-.10	.01	-.04	.04	-.04	.03																				
8. Conscientiousness	.01*/	-.11	.07*/	.08	.02	.02	.16/**																			
9. Extraversion	.13**/	-.03	.31**/**	.05	-.11*/	.07	.13/*	.22**/**																		
10. Neuroticism	.05	-.02	-.16*/**	.01	-.11	-.21**/**	-.22**/**	-.13*/	-.25**/**																	
11. Openness	.13/*	.01	.10	.03	.03	.10/**	.00	.07	.17**/	.02																
12. Self-esteem	.09	.06	.18*/**	.01	.10	.19**/*	.26**/**	.27**/**	.34**/**	-.58**/**	.05															
13. Self-monitoring	.02	.04	-.02	.06	.14*/	.13	.05	.13/**	.14**/	.05	.28*/**	-.05														
14. Dominance	.17**/*	.01	.34**/**	.04	-.02	.15**/	-.13/*	.16**/*	.51**/**	-.31**/**	.18*/**	.35**/**	.05													
15. Risk-taking tendency	.11	.09	.15*/	.03	.07	.17**/	-.17*/**	-.23**/**	.13*/	-.19**/*	.09	.10*/	.00	.29**/*												
16. General self-efficacy	.08	.02	.18*/**	.06	.07	.20**/**	.11*/	.20**/*	.35**/**	-.55**/**	.19*/*	.59**/**	.14/**	.45**/*	.26**/**											
17. Narcissism	.14*/	.00	.26**/**	-.01	-.02	.19**/	-.24**/**	.09*/	.46**/**	-.28**/**	.18*/*	.36**/**	.08	.65**/*	.36**/**	.47**/**										
18. *n* power	.00	.00	-.02	.03	-.02	-.08	-.06	-.03	.01	.02	-.03	.01	-.02	-.01	.03	-.02	.03									
19. Dominance motive	.11*/	-.02	.17*/	.02	.03	.14**/	-.51**/**	.13/**	.13/**	-.01	.05	.01	-.02	.42**/*	.33**/**	.17*/**	.57*/**	.08*/								
20. Leadership motive	.17**/*	.01	.35**/**	.06	.03	.18**/	-.15*/*	.15**/	.49**/**	-.27**/**	.12*/	.28**/**	.03	.73**/*	.24**/**	.39**/**	.63*/**	.01	.51**/**							
21. Prestige motive	.06	-.02	.08	-.03	.04	.07	-.03	.02	.16/**	.13*/	.05	.02	.02	.18*/*	.05	.05	.34*/**	.05	.34**/**	.29**/**						
22. Body height	.12**/	.07	.17**/	.03	.27**/**	.15**/	-.11/*	-.16*/**	-.04	-.20**/**	-.06	.04	-.11	.17*/*	.26**/**	.06	.13/*	.05	.16*/**	.16*/*	-.03					
23. Waist-to-hip-ratio	.07	.10	.08*/	-.03	.08/*	.15*/*	-.15/**	.21**/**	-.06	-.19**/**	-.10	-.01	-.15**/	.16*/*	.25**/**	.11/*	.11	.10	.19*/**	.13*/	-.04	.44**/**				
24. Testosterone[a]	.06	.07	.07	.01	.16/**	.17**/*	-.07	-.18**/**	-.14/*	-.28**/**	-.06	.06	-.13	.12	.26**/**	.10/*	.14*/*	.02	.18*/**	.13**/	-0.5	.61**/**	.59**/**			
25. Perceived attractiveness	-.09/**	-.06*/	.02	.03	.11/*	-.06	.13/*	.17*/*	.17**/**	-.01	.06	.09	.16*/*	.02	.01	-.01	.08	-.09	-.05	.02	.06	-.24**/**	-.33**/**	-.25**/**		
26. Perceived competence	.03	.03	.07	-.02	.01	.01	.11	.03	-.02	.05	.09	.06	.04	-.06	-.01	-.06	-.06	-.07	-.10	-.04	.01	.03	-.09/*	-.03	.38**/**	
27. Perceived trustworthiness	-.05	-.02	.01	-.01	-.05	-.09	.17**/*	.09	.06	.08	.07	.07	.10	-.11	-.08	-.04	-.10*/	-.05	-.15*/	-.08	.03	-.19**/**	-.26**/**	-.25**/**	.50**/**	.73**/

Note. GLI = General Leadership Impression. Because of missing values, sample sizes ranged from $N = 386$ to $N = 392$ in Study 1 and from $N = 376$ to $N = 380$ in Study. For success ratio, the sample size is $N = 332$ in Study 1 and $N = 299$ in Study 2. [a]All correlations using Testosterone are spearman correlations.

** $p < .001$. * $p < .01$. Significance Study 1/Significance Study 2

Corrections for Multiple Testing

Facing the problem of inflation of type I errors due to performing multiple hypothesis tests (i.e., performing one significance test for each predictor in a regression model), we applied the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995) to all central analyses using the predictor model. This procedure controls for the false discovery rate (FDR), that is, the expected proportion of significant results that are indeed false positives. The Benjamini-Hochberg correction offers the advantage of a powerful way of controlling the significance level without being overly conservative as, for example, the Bonferroni correction. In comparison, therefore, the Benjamini-Hochberg correction reduces the probability of rejecting effects that might exist (Diz, Carvajal-Rodríguez, & Skibinski, 2011). We applied an overall significance level of $\alpha = .05$, meaning that we expected no more than 5% of the significant tests to be false discoveries. In the following, we report uncorrected as well as Benjamini-Hochberg adjusted *p*-values (BH-adjusted *p*-values), thereby enabling the readers to draw their own conclusions.

Leadership Initiative

Descriptive statistics and initial checks. Figure 2 shows the distribution of leadership initiative in Study 1 and 2, respectively. In both studies, leadership initiative is distributed similarly and resembles a normal distribution. Additionally, we examined whether men or women showed more leadership initiative. To compare the amount of leadership initiative participants showed over both studies, we corrected for the different number of trials in the group task by dividing each participant's leadership initiative score by the number of trials her group worked on, resulting in the proportion of leadership initiative a participant showed over all trials. For example, a value of .30 would mean that in 30% of the

trials, a participant showed leadership initiative.[7] On average, men showed significantly more leadership initiative (M = 0.26, SD = 0.15) than women (M = 0.24, SD = 0.14), t(766.76) = 2.49, p = .013, d = 0.18. This rather small effect is probably due to the big sample size.

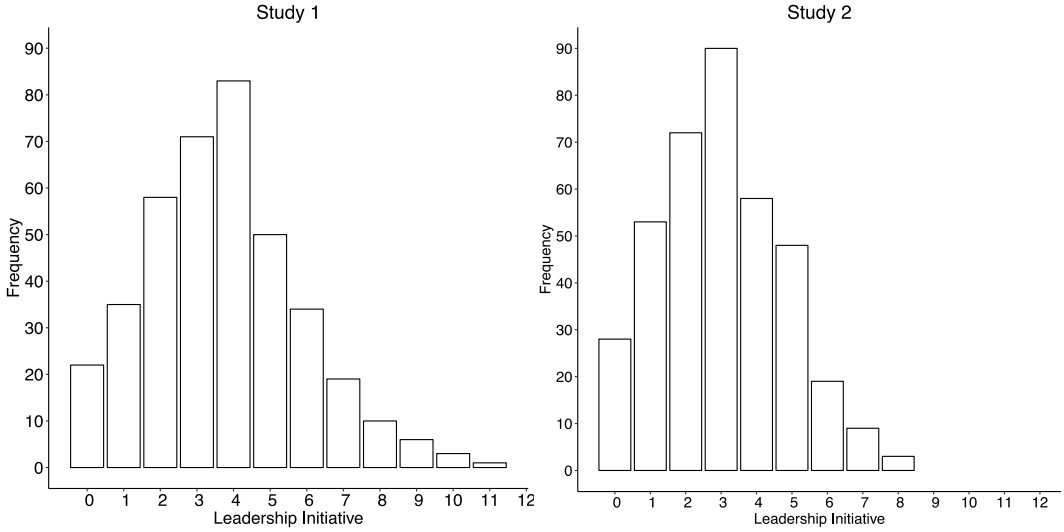

Figure 2. Distributions of Leadership Initiative for Study 1 and 2.

To check the validity of our operationalization of leadership initiative, we correlated leadership initiative with the traditionally used measure of leadership emergence, the General Leadership Impression Scale (GLI; Cronshaw & Lord, 1987). Capturing leader impressions of each group member, this correlation serves as an indicator of convergent validity. The two measures correlated significantly in both studies, r = .78 in Study 1, and r = .66 in Study 2, both ps < .001, indicating a substantial convergent validity.

Additionally, we checked whether leadership initiative is randomly distributed over the members of a group. The aim of this analysis is to ensure that the empirical distribution of leadership initiative does not equal a random distribution. This is an important prerequisite

[7] As described earlier, we excluded some trials for a few groups in Study 1 and applied a respective correction (i.e., dividing each leadership initiative score by the number of trials included in the analyses for that group).

because if it was random, how often a group member shows leadership initiative, predicting who is going to take the lead would become somewhat moot. For this analysis, we compared the empirical distribution of leadership initiative with a simulated random distribution for each study separately. For the empirical distribution, we ranked the four members of each group in descending order of the amount of leadership initiative they showed, and we then calculated the average amount of leadership initiative per rank over all groups. For the simulated random distribution, we simulated random distributions of leadership initiative for groups of four persons. This approach basically follows the logic of simulating how 15 or 12 balls for Study 1 and Study 2, respectively, fall randomly into four bins (n = 10000 times), which are then ordered by the number of balls that they contain. Subsequently, we calculated the mean of each bin. To compare the empirical and simulated means per rank, we performed one-sided t-tests. If leadership emergence is not a random process, then the person who shows leadership initiative the most should show a higher amount of leadership initiative than would be expected at random. Because of the resulting multiple testing, we applied a Bonferroni-corrected significance level ($\alpha = 0.05/4 = .0125$) and accepted the empirical distribution to be different from the simulated distribution if we find at least one significant difference between the ranks. Results can be found in Table 4 and Table 5. In both Study 1 and 2, means per rank 1 and 4 differed significantly from the simulated means per rank, meaning that over all groups, the person showing leadership initiative the most did so significantly more than would be expected from a random process, whereas the person showing leadership initiative the least did so significantly less than expected from a random process. We concluded that the amount of leadership initiative per member within the groups

is not randomly distributed and, hence, has a systematic component that needs to be further explored.[8]

Table 4

Empirical and simulated means per rank of leadership initiative in Study 1

Rank	Empirical mean (SD)	Simulated mean	*t(97)*	*p*	*d*
1	6.39 (1.49)	5.82	3.77	< .001	0.38
2	4.23 (0.77)	4.23	0.11	.915	0.01
3	2.86 (0.92)	3.09	-2.51	.014	0.25
4	1.40 (0.93)	1.86	-4.97	< .001	0.5

Table 5

Empirical and simulated means per rank of leadership initiative in Study 2

Rank	Empirical mean (SD)	Simulated mean	*t(97)*	*p*	*d*
1	5.24 (1.04)	4.88	3.46	< .001	0.35
2	3.46 (0.68)	3.39	0.84	.405	0.10
3	2.27 (0.76)	2.40	-1.50	.136	0.15
4	1.02 (0.77)	1.34	-3.91	< .001	0.41

Predicting leadership initiative. To investigate which variables are predictive of leadership initiative across both studies, we aggregated data from both studies to conduct the integrated analyses. Generally, the data structure is hierarchical (i.e., individuals are nested in groups, and groups are nested in studies). Because there can, by definition, be no differences in leadership initiative at the group level (in each group we find leadership initiative in every

[8] Additionally, we checked this prerequisite by simulating the means per rank differently. We simulated random distributions of leadership initiative for as many groups as were tested in each study (i.e., 98 groups in Study 1 and 95 groups in Study 2) and then calculated the means per ranks. We repeated this process $n = 10000$ times. In this way, we obtained 10000 means for each rank. We then compared the distribution of these simulated means per rank with our thresholds, the empirical means per rank. The results support the first simulation approach by finding significant differences for all means per rank in both studies, all $ps < .0125$, and revealing the lowest p-values in both simulations for the means of rank 1 and 4 (results are not reported).

trial, but not for each member within the group), there is no inter-group variance for that variable. For that reason, we did not include the group as a factor in our integrated analyses.

Results for the multiple regression analysis predicting leadership initiative can be found in Table 6. Applying the BH-adjusted significance level, we found that *g* factor, β = 0.03 (*SE* = 0.01), *t* = 3.73, BH-adjusted *p* < .001, subjective competence, β = 0.03 (*SE* = 0.01), *t* = 3.67, BH-adjusted *p* < .001, conscientiousness, β = 0.03 (*SE* = 0.01), *t* = 3.29, BH-adjusted *p* = .008, and extraversion, β = 0.03 (*SE* = 0.01), *t* = 3.29, BH-adjusted *p* = .008, significantly predicted leadership initiative over both studies. All of these effects were in a positive direction. Therefore, the more intelligent, the more subjective competent in solving the respective task, the more conscientious, and the more extraverted a person was compared to the other members of her group, the more leadership initiative she showed. The model had a medium effect size of f^2 = .30, representing a medium effect size (Cohen, 1988) and compared to each other, these effects had an equal proportion in it. Only one of these four effects was moderated by study: The effect of conscientiousness, β = -0.05 (*SE* = 0.01), *t* = -4.30, BH-adjusted *p* < .001, was larger in Study 1 compared to Study 2. Both of these models had a medium effect size, that is f^2 = .30 for Study 1 and f^2 = .15 for Study 2. Multiple regression analyses on study level revealed that conscientiousness was a significant positive predictor for leadership initiative in Study 1, β = 0.03, (*SE* = 0.01), *t* = 3.43, BH-adjusted *p* = .005, whereas it was a significant negative predictor in Study 2, β = -0.02, (*SE* = 0.01), *t* = -2.69, BH-adjusted *p* = .048. Therefore, the more conscientious a person was compared to her group in Study 1, the more leadership initiative she showed, whereas, in Study 2, the relationship was reversed. Furthermore, for one of those predictors that did not receive a significant regression weight overall, we also found a significant interaction with study, namely waist-to-hip-ratio, β = -0.03 (*SE* = 0.01), *t* = -3.02, BH-adjusted *p* = .021. Multiple regression analyses on study level revealed that in Study 1, the effect of waist-to-hip-ratio

was descriptively positive, but failed to reach significance, β = 0.02, (*SE* = 0.01), *t* = 2.49, BH-adjusted *p* = .052, whereas it was descriptively negative and also failed to reach significance in Study 2, β = -0.02, (*SE* = 0.01), *t* = -1.81, BH-adjusted *p* = .250. Therefore, descriptively, the higher the waist-to-hip-ratio of a person was in Study 1, the more leadership initiative she showed, whereas, in Study 2, the relationship was reversed. It is important to note that the effects on the study level were not significant. The interaction effect between perceived attractiveness and study failed to reach significance under the BH-adjusted *p* < .05 but reached significance under the uncorrected *p* < .05. We still qualify it for exploratory purposes in the following. Multiple regression analyses on study level revealed that attractiveness was not a significant predictor in Study 1, β = 0.00, (*SE* = 0.01), *t* = 0.05, BH-adjusted *p* = .962, whereas it was a significant predictor in Study 2, β = -0.03, (*SE* = 0.01), *t* = -3.25, BH-adjusted *p* = .008. In Study 2, the effect of attractiveness was in a negative direction. Therefore, in Study 2, the less attractive a participant was compared to her group, the more leadership initiative she showed. All other investigated effects failed to reach significance in this analysis.[9] [10]

---

[9] For exploratory purposes, we investigated the effect of several additional predictors for each study separately. To avoid multicollinearity, we replaced specific variables in our predictor model with the respective exploratory variables. We did not find new significant predictors (at a significance level of *p* < .01) when replacing body height with body weight, replacing the leadership, dominance, and prestige motive with the explicit power motive, replacing waist-to-hip-ratio with shoulder width, replacing the narcissism total score (measured by the NPI) by the NPI subscales leadership/authority, grandiose exhibitionism, and entitlement/exploitativeness (subscale calculations according to Ackerman et al., 2011), replacing the narcissism total score with the NARQ's narcissism score (Back et al., 2013) as well as with the NARQ's subscales admiration and rivalry (results of these models can be found in the online supplementary material).

[10] To explore the effect of our predictor model on leadership initiative further, we aimed at investigating which predictor variables were predictive for extreme groups of leadership initiative, that is, participants showing an extreme low (category 0) or extreme high amount of leadership initiative (category 1). This analysis might allow us to explore whether different predictor variables are predictive of these extreme groups, compared to the analysis including all participants. For the low leader category, we chose participants showing leadership initiative never or once in both studies (14.54% in Study 1; 21.32% in Study 2). For the high leader category, we chose participants showing leadership initiative at least six times in Study 1 (18.62%), or five times in Study 2 (20.79%). Unfortunately, we did not obtain usable results, because the model did not differ significantly from the intercept-only model that contains no predictors at all, $\chi^2(47) = 33.08$, $p = .938$. This is probably due to the relatively low number of observations included in this analysis in relation to the great number of predictors. Results for predicting extremes using logistic regression analyses can be found in the online supplementary material.

Table 6

Multiple regression analyses predicting leadership initiative over both studies

Predictor	*β*	*SE*	*t*(706)	*p*	BH-adjusted *p*
Intercept	0.25	0.01	36.82	< .001	< .001
g factor	0.03	0.01	3.73	< .001	< .001
Subjective competence	0.03	0.01	3.67	< .001	< .001
Agreeableness	-0.01	0.01	-1.47	.142	.487
Conscientiousness	0.03	0.01	3.29	.001	.008
Extraversion	0.03	0.01	3.29	.001	.008
Neuroticism	0.02	0.01	1.62	.105	.388
Openness	0.01	0.01	1.66	.097	.388
Self-esteem	0.01	0.01	0.53	.598	.844
Self-monitoring	0.00	0.01	-0.58	.562	.843
Dominance	0.01	0.01	0.76	.449	.843
Risk-taking tendency	0.01	0.01	1.31	.191	.536
General self-efficacy	-0.01	0.01	-1.33	.182	.536
Narcissism	0.00	0.01	0.26	.797	.964
n power	0.00	0.02	0.13	.896	.964
Dominance motive	0.00	0.01	0.40	.687	.916
Leadership motive	0.00	0.01	-0.33	.738	.957
Prestige motive	0.00	0.01	-0.05	.963	.964
Body height	0.02	0.01	1.83	.068	.301
Waist-to-hip-ratio	0.02	0.01	2.39	.017	.091
Testosterone	0.01	0.01	1.20	.230	.581
Perceived attractiveness	0.00	0.01	0.05	.964	.964
Perceived competence	0.01	0.01	1.14	.254	.610
Perceived trustworthiness	0.00	0.01	0.08	.936	.964

(continued on next page)

Table 6 (continued)

Predictor	*β*	*SE*	*t*(706)	*p*	BH-adjusted *p*
Study[a]	0.00	0.01	0.11	.911	0.954
Study × *g* factor	-0.01	0.01	-1.28	.201	.536
Study × Subjective competence	0.00	0.01	0.42	.677	.916
Study × Agreeableness	0.00	0.01	-0.06	.955	.964
Study × Conscientiousness	-0.05	0.01	-4.30	< .001	< .001
Study × Extraversion	-0.02	0.01	-1.43	.154	.493
Study × Neuroticism	-0.01	0.01	-0.98	.327	.747
Study × Openness	0.00	0.01	0.16	.870	.964
Study × Self-esteem	0.01	0.01	0.78	.439	.843
Study × Self-monitoring	0.01	0.01	0.60	.550	.843
Study × Dominance	0.01	0.02	0.59	.556	.843
Study × Risk-taking tendency	-0.01	0.01	-0.64	.522	.843
Study × General self-efficacy	0.00	0.01	0.15	.884	.964
Study × Narcissism	-0.01	0.02	-0.77	.444	.843
Study × *n* power	0.00	0.01	-0.07	.945	.964
Study × Dominance motive	-0.01	0.01	-0.60	.552	.843
Study × Leadership motive	0.01	0.02	0.83	.406	.843
Study × Prestige motive	0.01	0.01	0.53	.598	.844
Study × Body height	-0.02	0.01	-1.82	.069	.301
Study × Waist-to-hip-ratio	-0.03	0.01	-3.02	.003	.021
Study × Testosterone	-0.01	0.01	-0.62	.536	.843
Study × Perceived attractiveness	-0.03	0.01	-2.54	.011	.066
Study × Perceived competence	0.00	0.02	0.24	.808	.964
Study × Perceived trustworthiness	-0.01	0.02	-0.65	.518	.843

Note. R^2 = .23, adjusted R^2 = .18. $F(47, 706) = 4.51$, $p < .001$. N = 754. BH-adjusted p = Benjamini-Hochberg adjusted p-value. All predictors were centered on their group means and z-standardized.
[a]Study: 0 = Study 1, 1 = Study 2.

Robustness checks. Because leadership initiative is dependent within a group (i.e., if one group member shows leadership initiative very often, the other group members might show it rather rarely), we performed conditional logit analyses as a robust alternative. The conditional logit model is a special case of the logistic regression, which accounts for the

described dependency in our data. Although the conditional logit model offers this advantage, it does not allow for modeling effects of between-subject factors (in our case study is such a factor) and therefore is not suitable for an integrated analysis. Therefore, we performed separate analyses for each study using the conditional logit model. Because we performed these analyses on the study level, we used an adjusted predictor model without any interaction terms.

Applied to leadership initiative, the conditional logit model predicted whether a person in a group showed leadership initiative in a trial (coded with 1) or did not show leadership initiative (coded with 0). Results for the conditional logit analysis for Study 1 can be found in Table 7. We found significant main effects for *g* factor, $\beta = 0.15$ (*SE* = 0.03), $z = 4.40$, *OR* = 1.16, BH-adjusted $p < .001$, subjective competence, $\beta = 0.16$ (*SE* = 0.04), $z = 4.40$, *OR* = 1.17, BH-adjusted $p < .001$, conscientiousness, $\beta = 0.14$ (*SE* = 0.04), $z = 3.73$, *OR* = 1.15, BH-adjusted $p < .001$, extraversion, $\beta = 0.15$ (*SE* = 0.04), $z = 3.67$, *OR* = 1.17, BH-adjusted $p < .001$, and waist-to-hip-ratio, $\beta = 0.10$ (*SE* = 0.04), $z = 2.79$, *OR* = 1.11, BH-adjusted $p = .023$. All of these effects were in a positive direction. Therefore, the more intelligent, the more subjective competent in solving the respective task, the more conscientious, the more extraverted, and the higher the waist-to-hip-ratio of a person was compared to her group members, the more likely it was that she showed leadership initiative. Relative to each other, the odds that a person showed leadership initiative in a trial compared to she did not show leadership initiative increased by 1.17, and therefore the most, the more extraverted and subjective competent a person was, followed by an increase by 1.16 the more intelligent a person was, and an increase by 1.15 the more conscientious a person was compared to her group. Among the significant effects, the increase in odds by 1.11 was lowest for waist-to-hip-ratio. No other tested effects reached significance.

Table 7

Conditional logit analyses predicting leadership initiative in Study 1

Predictor	β	*SE*	*OR*	*z*	*p*	BH-adjusted *p*
g factor	0.15	0.03	1.16	4.40	< .001	< .001
Subjective competency	0.16	0.04	1.17	4.43	< .001	< .001
Agreeableness	-0.08	0.04	0.92	-1.89	.058	.164
Conscientiousness	0.14	0.04	1.15	3.73	< .001	< .001
Extraversion	0.15	0.04	1.17	3.67	< .001	< .001
Neuroticism	0.08	0.04	1.08	1.85	.064	.164
Openness	0.07	0.03	1.07	1.89	.059	.164
Self-esteem	0.03	0.05	1.03	0.67	.506	.727
Self-monitoring	-0.03	0.04	0.97	-0.73	.468	.718
Dominance	0.04	0.05	1.05	0.85	.394	.647
Risk-taking tendency	0.05	0.04	1.05	1.45	.147	.282
General self-efficacy	-0.07	0.05	0.93	-1.52	.129	.270
Narcissism	0.02	0.05	1.02	0.34	.731	.877
n power	0.01	0.03	1.01	0.25	.801	.877
Dominance motive	0.02	0.05	1.02	0.47	.636	.860
Leadership motive	-0.01	0.05	0.99	-0.26	.793	.877
Prestige motive	-0.01	0.04	0.99	-0.30	.762	.877
Body height	0.09	0.04	1.09	2.22	.027	.104
Waist-to-hip-ratio	0.10	0.04	1.11	2.79	.005	.023
Testosterone	0.04	0.04	1.04	1.06	.289	.511
Perceived attractiveness	0.00	0.04	1.00	0.02	.986	.986
Perceived competence	0.07	0.05	1.01	1.55	.121	.270
Perceived trustworthiness	0.01	0.05	1.01	0.13	.900	.941

Note. R^2 = .03 (maximum possible R^2 = .65). Wald test $\chi^2(23) = 160.90$, $p < .001$. $N = 5832$. BH-adjusted *p* = Benjamini-Hochberg adjusted *p*-value. All predictors were centered on their group means and z-standardized.

Table 8 shows the results of the conditional logit analyses predicting leadership initiative in Study 2. We found significant main effects for subjective competence, $\beta = 0.17$ (*SE* = 0.04), *z* = 4.50, *OR* = 1.19, BH-adjusted *p* < .001, conscientiousness, $\beta = -0.12$ (*SE* = 0.4), *z* = -2.85, *OR* = 0.89, BH-adjusted *p* = .031, and perceived attractiveness, $\beta = -0.17$ (*SE*

= 0.05), z = -3.55, OR = 0.84, BH-adjusted $p < .001$. The effect of subjective competence was in a positive direction, whereas the effects of conscientiousness and perceived attractiveness were negative. Therefore, the more conscientious a person was compared to her group, the more likely it was that she showed leadership initiative in a trial. Relative to each other, the change in odds for showing leadership initiative was greatest for the effects of subjective competence and perceived attractiveness. The odds that a person showed leadership initiative compared to she did not show leadership initiative in a trial increased by 1.19, the more subjective competent a person was compared to her group. The odds that a person showed leadership initiative compared to she did not show leadership initiative were 0.89 times smaller if a person was more attractive compared to her group, followed by being 0.84 times smaller if a person was more conscientious compared to her group. All other tested effects did not reach significance.

Table 8

Conditional logit analyses predicting leadership initiative in Study 2

Predictor	*β*	*SE*	*OR*	*z*	*p*	BH-adjusted *p*
g factor	0.08	0.04	1.08	2.12	.034	.161
Subjective competency	0.17	0.04	1.19	4.50	< .001	< .001
Agreeableness	-0.08	0.05	0.92	-1.78	.075	.216
Conscientiousness	-0.12	0.04	0.89	-2.85	.004	.031
Extraversion	0.06	0.05	1.06	1.25	.213	.445
Neuroticism	0.01	0.05	1.01	0.16	.872	.912
Openness	0.08	0.04	1.09	2.11	.035	.161
Self-esteem	0.09	0.05	1.10	1.82	.068	.216
Self-monitoring	0.01	0.04	1.01	0.29	.776	.850
Dominance	0.10	0.06	1.10	1.62	.105	.268
Risk-taking tendency	0.02	0.04	1.02	0.40	.688	.819
General self-efficacy	-0.06	0.06	0.94	-1.13	.259	.496
Narcissism	-0.06	0.06	0.94	-0.98	.328	.567
n power	0.00	0.04	1.00	-0.05	.962	.962
Dominance motive	-0.02	0.05	0.98	-0.40	.687	.819
Leadership motive	0.05	0.06	1.05	0.77	.442	.624
Prestige motive	0.04	0.04	1.04	0.91	.362	.567
Body height	-0.03	0.04	0.97	-0.74	.461	.624
Waist-to-hip-ratio	-0.08	0.04	0.92	-1.94	.052	.199
Testosterone	0.02	0.04	1.02	0.37	.712	.819
Perceived attractiveness	-0.17	0.05	0.84	-3.55	< .001	< .001
Perceived competence	0.08	0.06	1.09	1.50	.134	.308
Perceived trustworthiness	-0.05	0.06	0.95	-0.90	.370	.567

Note. R^2 = .02 (maximum possible R^2 = .65). Wald test $\chi^2(23) = 95.40$, $p < .001$. $N = 4560$. BH-adjusted *p* = Benjamini-Hochberg adjusted *p*-value. All predictors were centered on their group means and z-standardized.

In summary, the results of both conditional logit analyses on study level reflected the pattern of results we found in the integrated analysis for leadership initiative across both studies. For example, we found a significant main effect of subjective competence in the integrated analysis, and also the main effect for subjective competence in each study. Another

example is the effect of conscientiousness. We found a positive main effect of conscientiousness in the integrated analysis, that is also moderated by study. In the conditional logit analyses, we find a significant positive main effect of conscientiousness in Study 1 and a slightly weaker negative significant effect in Study 2. In a similar way, the other found effects complement the pattern of results we found in the integrated analysis across both studies.

Group gender composition. For exploratory purposes, we examined whether the effects of the predictor variables on our criterion variables vary according to the groups' gender composition. To avoid overfitting, we aimed at avoiding modeling interaction effects between every predictor from the predictor model and the factor group gender composition. Therefore, we first selected predictor variables with significant main effects under the BH-adjusted significance level $\alpha < .10$ for the respective criterion variable. Subsequently, we modeled their interaction effects with group gender composition (mixed-gender, male, female). We established this moderate criterion for predictor inclusion in order to be able to address effects in the group gender composition analyses that might have been canceled out in the analyses, including all groups. As a hypothetical example, consider the effect of agreeableness on leadership initiative was significant in female groups, and significant, but negative, for male groups, whereas there was no considerable effect in mixed-gender groups. In an analysis over all groups, the effects of agreeableness could have canceled out or barely missed the BH-adjusted significance level. For simplicity, we only modeled interactions between the predictors and group composition, thereby preventing three-way-interactions with study in our model. We will not report the main effects in the following, because we only used a part of the predictor model in these analyses and could possibly find other effects than in the respective main analyses. Therefore, we only focused on interaction effects. To test the interactions, we dummy-coded interaction contrasts using mixed-gender groups as the

baseline category, comparing them with female and male groups. The online supplementary material includes results for the corresponding analyses for each separate study. For these analyses, we applied the same set of predictors for each criterion variable that we used in the following integrated analyses.

To investigate the possible moderating effect of group gender composition on effects of predictor variables on leadership initiative, we included *g* factor, subjective competence, and extraversion as predictor variables. Results can be found in Table 9. The model had a medium effect size of f^2 = .16. Under the BH-adjusted significance level, we found a significant interaction contrast for subjective competence, comparing its effect between mixed-gender and male groups, β = 0.05 (*SE* = 0.01), *t* = 3.78, BH-adjusted *p* < .001, indicating that the relation between subjective competence and leadership initiative differed between these two groups. Subgroup analyses revealed that subjective competence was not a significant predictor for leadership initiative in mixed-gender groups, β = 0.01 (*SE* = 0.01), *t* = 1.45, BH-adjusted *p* = .178, whereas it significantly predicted leadership initiative in male groups, β = 0.06 (*SE* = 0.01), *t* = 7.02, BH-adjusted *p* < .001. In female groups, subjective competence also significantly predicted leadership initiative, β = 0.02 (*SE* = 0.01), *t* = 2.81, BH-adjusted *p* = .010, but to a lesser extent than in male groups. All models for subgroup analyses had medium effect sizes, that is, f^2 = .14 for the model for mixed-gender groups, f^2 = .23 in male groups, and f^2 = .12 for female groups. Therefore, compared to their groups, more subjective competent participants showed more leadership initiative in male and female groups, whereby that relation was strongest in male groups. In mixed-gender groups, more subjective competent participants did not show more leadership initiative. The interactions between waist-to-hip-ratio and both interaction contrasts, comparing its effect between female and mixed-gender groups, β = -0.03 (*SE* = 0.01), *t* = -2.41, *p* = .016, BH-adjusted *p* = .058, and comparing its effect between male and mixed-gender groups, β = -0.03 (*SE* = 0.01),

t = -2.24, p = .025, BH-adjusted p = .075, failed to reach significance under the BH-adjusted $p < .05$, but reached significance under the uncorrected $p < .05$. For exploratory purposes, we qualify these effects in the following. Subgroup analyses revealed that waist-to-hip-ratio was a significant predictor for leadership initiative in mixed-gender groups under both p-values, β = 0.03 (SE = 0.01), t = 3.72, $p < .001$, BH-adjusted $p < .001$, but was not a significant predictor in female groups, β = -0.01 (SE = 0.01), t = -0.53, p = 0.60, BH-adjusted p = .596, or male groups, β = 0.00 (SE = 0.01), t = -0.26, p = 0.794, BH-adjusted p = .794. Therefore, the higher a participant's waist-to-hip-ratio was compared to her group, the more leadership she showed in mixed-gender groups, and that relation was stronger as in male or female groups. All other interaction effects in this analysis did not reach significance.

Table 9

Multiple regression analyses predicting the effects of group gender composition on leadership initiative

Predictor	*β*	*SE*	*t*(751)	*p*	BH-adjusted *p*
Intercept	0.25	0.01	29.72	< .001	< .001
g factor	0.03	0.01	3.66	< .001	< .001
Subjective competence	0.01	0.01	1.52	.130	.260
Conscientiousness	-0.01	0.01	-0.77	.440	.660
Extraversion	0.02	0.01	2.18	.030	.077
Waist-to-hip-ratio	0.03	0.01	3.88	< .001	< .001
GGC[a]: f vs. b × *g* factor	0.00	0.01	-0.37	.708	.910
GGC[a]: m vs. b × *g* factor	-0.02	0.01	-1.57	.117	.260
GGC[a]: f vs. b × Subjective competence	0.01	0.01	0.82	.415	.660
GGC[a]: m vs. b × Subjective competence	0.05	0.01	3.78	< .001	< .001
GGC[a]: f vs. b × Conscientiousness	0.02	0.01	1.32	.188	.338
GGC[a]: m vs. b × Conscientiousness	-0.01	0.01	-0.67	.504	.698
GGC[a]: f vs. b × Extraversion	0.00	0.01	-0.09	.930	.985
GGC[a]: m vs. b × Extraversion	0.00	0.01	0.14	.892	.985
GGC[a]: f vs. b × Waist-to-hip-ratio	-0.03	0.01	-2.41	.016	.058
GGC[a]: m vs. b × Waist-to-hip-ratio	-0.03	0.01	-2.24	.025	.075

Note. R^2 = .16, adjusted R^2 =.14, $F(17, 751) = 8.19$, $p < .001$, $N = 769$. BH-adjusted *p* = Benjamini-Hochberg adjusted *p*-value. GGC = Group gender composition. All predictors were centered on their group means and z-standardized.

[a]The factor group gender composition was dummy-coded with mixed-gender groups as the baseline (b) category, compared to either female (f) or male groups (m).

Success of Leadership Initiative

Descriptive statistics and initial checks. Over both studies, 77.29% of the proposed answers were ultimately accepted as the group's decision. There was no significant association between participants' gender and success of leadership initiative. Men and women were equally successful in their leadership attempts, $\chi^2(1, N = 2510) = 1.46$, $p = 0.23$. Men were successful in 72.17% of their leadership attempts, whereas women were successful in 68.77% of their leadership attempts.

Predicting success of leadership initiative. To examine which variables are predictive of success of leadership initiative, we used all variables that were part of our predictor model for leadership initiative, while also adding two additional variables: It is plausible that a participant's success in attracting followers might be associated with the mean accuracy of the answers she already proposed. It would be a highly rational strategy of the group members to agree to accurate answers. Similarly, a participant's success of leadership initiative might be due to the number of answers she already proposed, that is, leadership initiative. Participants might have inferred that a person, who already proposed several answers, has good reasons to do so and therefore are more likely to agree to that person's proposed answers. That is why we added the main effects as well as the interaction effects of these two variables with the factor study to the predictor model. It is important to note that, in contrast to other predictors, a significant regression weight of leadership initiative when predicting success of leadership initiative would not directly indicate a causal influence of leadership initiative, because reverse causality (i.e., experiencing success increases the likelihood of trying it again) is also possible.

Results for the logistic regression analysis predicting success of leadership initiative can be found in Table 10. Under the BH-adjusted significance level, we found significant main effects for *g* factor, $\beta = 0.20$ (*SE* = 0.07), $z = 2.94$, *OR* = 1.23, BH-adjusted $p = .031$, accuracy of the proposed answers, $\beta = 0.21$ (*SE* = 0.08), $z = 3.04$, *OR* = 1.24, BH-adjusted $p = .026$, and leadership initiative, $\beta = 0.26$ (*SE* = 0.11), $z = 3.50$, *OR* = 1.30, BH-adjusted $p < .001$. All of these effects were in a positive direction. That means that the more intelligent a participant was compared to her group, the more likely it was that the group accepted her proposed answer. Similarly, the higher she assessed her competence in solving the respective tasks, and the more answers she proposed compared to her group, the more likely it was that the group accepted her proposed answer. Relative to each other, the increase in odds that a

suggested answer was accepted compared to it was rejected increased most, by 1.30, for leadership initiative, followed by an increase by 1.24 for the accuracy of the proposed answers and by 1.23 for *g* factor. We did not find any significant interaction effects involving the study factor so that none of these effects was moderated by study. In this analysis, all other investigated effects failed to reach significance.[11]

[11] In the same way we investigated the effects of exploratory variables on leadership initiative, we tested their effects on success of leadership initiative. We did not find any new significant predictor variables (again, results of these models can be found in the online supplementary material).

Table 10

Binary logistic regression analyses predicting success of leadership initiative

Predictor	*β*	*SE*	*OR*	*z*	*p*	BH-adjusted *p*
Intercept	0.88	0.07	2.41	12.10	< .001	< .001
g factor	0.20	0.07	1.23	2.94	.003	.031
Subjective competence	-0.07	0.07	0.93	-0.96	.337	.743
Agreeableness	0.04	0.08	1.04	0.45	.655	.921
Conscientiousness	-0.15	0.08	0.86	-1.94	.052	.276
Extraversion	-0.01	0.09	0.99	-0.08	.935	.976
Neuroticism	0.01	0.09	1.01	0.17	.869	.976
Openness	-0.06	0.07	0.94	-0.85	.396	.743
Self-esteem	0.00	0.09	1.00	0.03	.980	.992
Self-monitoring	0.12	0.07	1.13	1.66	.098	.463
Dominance	-0.03	0.10	0.97	-0.31	.753	.923
Risk-taking tendency	0.12	0.07	1.12	1.61	.107	.464
General self-efficacy	-0.06	0.09	0.94	-0.70	.483	.838
Narcissism	-0.04	0.11	0.96	-0.40	.692	.923
n power	0.00	0.06	1.00	0.01	.992	.992
Dominance motive	-0.04	0.10	0.96	-0.37	.713	.923
Leadership motive	-0.03	0.07	0.97	-0.33	.744	.923
Prestige motive	0.04	0.08	1.04	0.57	.566	.866
Body height	-0.10	0.07	0.91	-1.24	.213	.691
Waist-to-hip-ratio	-0.05	0.08	0.95	-0.65	.513	.838
Testosterone	0.13	0.10	1.13	1.53	.126	.504
Perceived attractiveness	-0.08	0.07	0.92	-1.02	.308	.743
Perceived competence	0.06	0.10	1.06	0.66	.512	.838
Perceived trustworthiness	-0.10	0.08	0.91	-0.99	.323	.743
Accuracy	0.21	0.08	1.24	3.04	.002	.026
Leadership initiative	0.26	0.11	1.30	3.50	< .001	< .001

(continued on next page)

Table 10 (continued)

Predictor	*β*	*SE*	*OR*	*z*	*p*	BH-adjusted *p*
Study[a]	0.62	0.13	1.86	4.88	< .001	< .001
Study × *g* factor	-0.01	0.12	0.99	-0.13	.899	.976
Study × Subjective competence	0.12	0.14	1.12	0.99	.320	.743
Study × Agreeableness	-0.17	0.12	0.85	-1.21	.226	.691
Study × Conscientiousness	0.07	0.15	1.07	0.54	.592	.880
Study × Extraversion	0.12	0.15	1.13	0.84	.400	.743
Study × Neuroticism	-0.17	0.15	0.85	-1.12	.263	.743
Study × Openness	0.17	0.12	1.19	1.48	.140	.520
Study × Self-esteem	-0.14	0.16	0.87	-0.87	.383	.743
Study × Self-monitoring	-0.12	0.12	0.89	-0.95	.345	.743
Study × Dominance	0.05	0.18	1.06	0.30	.763	.923
Study × Risk-taking tendency	-0.01	0.13	0.99	-0.10	.918	.976
Study × General self-efficacy	0.21	0.17	1.23	1.25	.210	.691
Study × Narcissism	-0.11	0.18	0.89	-0.63	.532	.838
Study × *n* power	0.03	0.11	1.03	0.23	.817	.966
Study × Dominance motive	-0.07	0.16	0.93	-0.46	.649	.921
Study × Leadership motive	0.07	0.18	1.07	0.37	.709	.923
Study × Prestige motive	0.01	0.13	1.01	0.11	.909	.976
Study × Body height	0.12	0.13	1.12	0.91	.361	.743
Study × Waist-to-hip-ratio	0.25	0.13	1.28	1.93	.053	.276
Study × Testosterone	-0.33	0.13	0.72	-2.48	.013	.113
Study × Perceived attractiveness	0.32	0.14	1.37	2.22	.026	.193
Study × Perceived competence	0.18	0.17	1.19	1.07	.286	.743
Study × Perceived trustworthiness	-0.11	0.17	0.90	-0.62	.532	.838
Study × Accuracy	-0.24	0.12	0.79	-1.99	.047	.276
Study × Leadership initiative	-0.01	0.13	0.99	-0.08	.938	.976

Note. Hosmer and Lemeshow R^2 = .05, Cox and Snell R^2 = .05, Nagelkerke R^2 = .08. $\chi^2(51) = 129.92$, $p < .001$. $N = 2460$. BH-adjusted p = Benjamini-Hochberg adjusted p-value. All predictors were centered on their group means and z-standardized.

[a]Study: 0 = Study 1, 1 = Study 2.

Group gender composition. In order to explore the effect of group gender composition on success of leadership initiative, we applied the same approach as with leadership initiative. In line with the moderate criterion for predictor inclusion, we tested whether the associations of general intelligence, accuracy, and leadership initiative with success of leadership initiative differed between the three group gender compositions. Table 11 shows the results. Under the BH-adjusted significance level, we did not find any significant interaction contrasts, all $\beta s \leq 0.21$ ($SEs \leq 0.14$), $ORs \leq 1.23$, $zs \leq 1.54$, BH-adjusted $ps \geq .284$. The effects of general intelligence, accuracy, and leadership initiative did not differ between the group gender compositions.

Table 11

Binary logistic regression analyses predicting the effects of group gender composition on success of leadership initiative

Predictor	*β*	*SE*	*OR*	*z*	*p*	BH-adjusted *p*
Intercept	1.11	0.10	3.03	11.22	<.001	<.001
g factor	0.28	0.09	1.32	3.18	.001	.006
Accuracy	0.10	0.09	1.12	1.14	.256	.439
Leadership initiative	0.15	0.09	1.16	1.66	.098	.284
GGC[a]: f vs. b	-0.08	0.14	0.92	-0.59	.555	.666
GGC[a]: m vs. b	-0.03	0.14	0.97	-0.19	.848	.848
GGC[a]: f vs. b × *g* factor	-0.20	0.12	0.82	-1.66	.097	.284
GGC[a]: m vs. b × *g* factor	-0.18	0.12	0.84	-1.47	.142	.284
GGC[a]: f vs. b × Accuracy	-0.05	0.12	0.95	-0.40	.696	.759
GGC[a]: m vs. b × Accuracy	0.13	0.14	1.14	0.97	.333	.500
GGC[a]: f vs. b × Leadership initiative	0.09	0.14	1.10	0.67	.501	.666
GGC[a]: m vs. b × Leadership initiative	0.21	0.13	1.23	1.54	.123	.284

Note. Hosmer and Lemeshow $R^2 = .02$, Cox and Snell $R^2 = .02$, Nagelkerke $R^2 = .03$. $\chi^2(11) = 44.79$, $p < .001$. $N = 2510$. BH-adjusted p = Benjamini-Hochberg adjusted p-value. GGC = Group gender composition. All predictors were centered on their group means and z-standardized.

[a]The factor group gender composition was dummy-coded with mixed-gender groups as the baseline (b) category, compared to either female (f) or male groups (m).

Explorative Analyses

Overall leadership.

Descriptive statistics. Over both studies, participants showed leadership initiative, and their answers were accepted as the group's decision in 19.18% of the cases. There was a significant association between participant's gender and overall leadership, $\chi^2(1, N = 10392) = 12.20$, $p < .001$. The odds of showing leadership initiative and getting the proposed answer accepted were 1.19 times lower if the participant was a woman compared to a man.

As for leadership initiative, we checked the validity of our operationalization of overall leadership. We correlated the total number of successful leadership attempts with the traditionally used measure of leadership emergence, the General Leadership Impression Scale (GLI; Cronshaw & Lord, 1987). The two measures correlated significantly in both studies, $r = .73$ in Study 1, and $r = .66$ in Study 2, both $ps < .001$, indicating a substantial convergent validity.

Predicting overall leadership. We predicted overall leadership using the predictor model and performing logistic regression analysis. Results can be found in Table 12. Under the BH-adjusted significance level, we found significant main effects for *g* factor, $\beta = 0.22$ ($SE = 0.04$), $z = 5.61$, $OR = 1.23$, BH-adjusted $p < .001$, subjective competence, $\beta = 0.14$ ($SE = 0.04$), $z = 3.48$, $OR = 1.15$, BH-adjusted $p = .008$, and extraversion $\beta = 0.15$ ($SE = 0.05$), $z = 3.26$, $OR = 1.16$, BH-adjusted $p = .008$. All of these effects were in a positive direction. The more intelligent a participant was, the better she assessed her subjective competence in solving the respective task, and the more extraverted a person was compared to her group, the more likely it was that she showed leadership initiative and got her proposed answer accepted. Relative to each other, the increase in odds that a person showed leadership initiative and was successful increased most, by 1.23, for *g* factor, followed by an increase by 1.16 for extraversion and by 1.15 for subjective task competence. None of the three

significant effects was moderated by study. Furthermore, for the predictor conscientiousness, that failed to reach a significant regression weight overall, we also found a significant interaction with study, β = -0.24 (*SE* = 0.06), *z* = -4.04, *OR* = 0.78, BH-adjusted $p < .001$. The effect of conscientiousness was moderated by study and was larger in Study 1 compared to Study 2. Logistic regression analyses on study level revealed that the effect of conscientiousness on overall leadership was significant in Study 1, β = 0.11 (*SE* = 0.04), *z* = 2.66, *OR* = 1.12, BH-adjusted $p = .038$, as well as in Study 2, β = -.013 (*SE* = 0.04), *z* = -3.04, *OR* = 0.87, BH-adjusted $p = .016$. It is important to note that the effect of conscientiousness was positive in Study 1 but negative in Study 2. Therefore, in Study 1, the more conscientious a person was, the more likely it was that she showed leadership initiative, and her answer was accepted, whereas, in Study 2, that relationship was reversed. In this analysis, all other investigated effects failed to reach significance. Moreover, we only had a few interaction effects between the predictors and study, indicating that our results were very independent of the study.

Table 12

Binary logistic regression analyses predicting overall leadership

Predictor	*β*	*SE*	*OR*	*z*	*p*	BH-adjusted *p*
Intercept	-1.57	0.04	0.21	-43.05	< .001	< .001
g factor	0.22	0.04	1.23	5.61	< .001	< .001
Subjective competence	0.14	0.04	1.15	3.48	.001	.008
Agreeableness	-0.08	0.05	0.92	-1.70	.089	.267
Conscientiousness	0.11	0.04	1.12	2.66	.008	.055
Extraversion	0.15	0.05	1.16	3.26	.001	.008
Neuroticism	0.10	0.05	1.10	1.96	.050	.218
Openness	0.06	0.04	1.06	1.53	.125	.316
Self-esteem	0.04	0.05	1.04	0.78	.436	.655
Self-monitoring	0.01	0.04	1.00	0.22	.825	.982
Dominance	0.02	0.06	1.02	0.37	.713	.925
Risk-taking tendency	0.10	0.04	1.10	2.46	.014	.084
General self-efficacy	-0.09	0.05	0.91	-1.83	.067	.267
Narcissism	0.00	0.06	1.00	-0.05	.964	.999
n power	0.00	0.04	1.00	0.01	.993	.999
Dominance motive	0.00	0.06	1.00	-0.06	.953	.999
Leadership motive	0.00	0.06	1.00	0.00	.999	.999
Prestige motive	0.01	0.04	1.01	0.16	.874	.999
Body height	0.06	0.04	1.06	1.35	.175	.400
Waist-to-hip-ratio	0.09	0.04	1.10	2.27	.023	.123
Testosterone	0.08	0.05	1.08	1.73	.084	.267
Perceived attractiveness	-0.04	0.04	0.96	-0.86	.391	.655
Perceived competence	0.09	0.05	1.09	1.60	.110	.293
Perceived trustworthiness	-0.01	0.06	0.99	-0.26	.792	.975

(continued on next page)

Table 12 (continued)

Predictor	*β*	*SE*	*OR*	*z*	*p*	BH-adjusted *p*
Study[a]	0.19	0.05	1.21	3.64	< .001	< .001
Study × *g* factor	-0.10	0.06	0.90	-1.78	.076	.267
Study × Subjective competence	0.04	0.06	1.05	0.78	.436	.655
Study × Agreeableness	-0.01	0.07	0.99	-0.20	.839	.982
Study × Conscientiousness	-0.24	0.06	0.78	-4.04	< .001	< .001
Study × Extraversion	-0.08	0.07	0.93	-1.12	.263	.549
Study × Neuroticism	-0.12	0.07	0.89	-1.62	.105	.293
Study × Openness	0.04	0.06	1.05	0.78	.438	.655
Study × Self-esteem	0.02	0.08	1.02	0.29	.773	.975
Study × Self-monitoring	-0.01	0.06	0.99	-0.12	.904	.999
Study × Dominance	0.07	0.09	1.07	0.80	.421	.655
Study × Risk-taking tendency	-0.06	0.06	0.94	-0.91	.361	.655
Study × General self-efficacy	0.06	0.08	1.06	0.76	.450	.655
Study × Narcissism	-0.08	0.09	0.93	-0.84	.400	.655
Study × *n* power	0.00	0.05	1.00	-0.04	.966	.999
Study × Dominance motive	-0.05	0.08	0.95	-0.61	.540	.762
Study × Leadership motive	0.07	0.09	1.07	0.82	.412	.655
Study × Prestige motive	0.03	0.06	1.03	0.43	.668	.891
Study × Body height	-0.09	0.06	0.91	-1.50	.135	.324
Study × Waist-to-hip-ratio	-0.13	0.06	0.88	-2.12	.034	.163
Study × Testosterone	-0.11	0.06	0.89	-1.74	.081	.267
Study × Perceived attractiveness	-0.08	0.07	0.92	-1.18	.238	.519
Study × Perceived competence	0.05	0.08	1.04	0.57	.571	.783
Study × Perceived trustworthiness	-0.09	0.08	0.92	-1.02	.310	.620

Note. Hosmer and Lemeshow R^2 = .03, Cox and Snell R^2 = .03, Nagelkerke R^2 = .04. $\chi^2(47) = 256.92$, $p < .001$, $N = 10134$. BH-adjusted p = Benjamini-Hochberg adjusted p-value. All predictors were centered on their group means and z-standardized.
[a]Study: 0 = Study 1, 1 = Study 2.

Group gender composition. To investigate the effect of group gender composition on overall leadership, we included *g* factor, subjective competence, conscientiousness, extraversion, and risk-taking tendency in the model. Results can be found in Table 13. Like for the group gender composition analyses for leadership initiative, we found a significant

interaction contrast for subjective competence, comparing its effect between mixed-gender and male groups, β = 0.19 (*SE* = 0.06), *z* = 3.04, *OR* = 1.21, BH-adjusted *p* = .009. Subgroup analysis showed that subjective competence just reached significance as a predictor for overall leadership in mixed-gender groups, β = 0.10 (*SE* = 0.05), *z* = 2.21, *OR* = 1.11, BH-adjusted *p* = .040. In male groups, subjective competence was as well a significant predictor for overall leadership, β = 0.29 (*SE* = 0.04), *z* = 6.62, *OR* = 1.34, BH-adjusted *p* < .001. Therefore, the effect of subjective competence on overall leadership was stronger in male groups. In female groups, subjective competence was also a significant predictor, β = 0.13 (*SE* = 0.05), *z* = 2.69, *OR* = 1.14, BH-adjusted *p* = .014, and its effect was weaker than in male groups, but stronger than in mixed-gender groups. Therefore, the more subjective competent a participant was compared to her group, the more likely it was that her leadership attempt was successful in male groups, and that relationship was weaker in female groups, followed by mixed-gender groups. We found a significant interaction contrast for risk-taking tendency, comparing its effect between mixed-gender and female groups, β = -0.18 (*SE* = 0.07), *z* = -2.58, *OR* = 0.84, BH-adjusted *p* = .036. Subgroup analyses revealed that risk-taking tendency was a significant predictor for overall leadership in mixed-gender groups, β = 0.17 (*SE* = 0.05), *z* = 3.50, *OR* = 1.18, BH-adjusted *p* < .001, whereas it was not a significant predictor for overall leadership in female groups, β = -0.01 (*SE* = 0.05), *z* = -0.28, *OR* = 0.99, BH-adjusted *p* = .797. In male groups, risk-taking tendency was not a significant predictor for overall leadership, β = 0.07 (*SE* = 0.04), *z* = 1.6, *OR* = 1.07, BH-adjusted *p* = .134. Therefore, the more a participant was prone to take risks compared to her group, the more likely it was that her leadership attempt was successful in mixed-gender groups. In male and female groups, it was not more likely that participants who were more prone to take risks were successful in their leadership attempts.

Table 13

Binary logistic regression analyses predicting the effects of group gender composition on overall leadership

Predictor	*β*	*SE*	*OR*	*z*	*p*	BH-adjusted *p*
Intercept	-1.47	0.04	0.23	-32.99	< .001	< .001
g factor	0.19	0.05	1.21	4.05	< .001	< .001
Subjective competence	0.10	0.05	1.10	2.21	.027	.081
Conscientiousness	-0.05	0.04	0.96	-1.02	.306	.501
Extraversion	0.04	0.05	1.04	0.76	.445	.616
Risk-taking tendency	0.17	0.05	1.18	3.50	< .001	< .001
GGC[a]: f vs. b	-0.01	0.06	0.99	-0.16	.872	.897
GGC[a]: m vs. b	0.01	0.06	1.01	0.20	.842	.897
GGC[a]: f vs. b × *g* factor	0.01	0.06	1.01	0.13	.897	.897
GGC[a]: m vs. b × *g* factor	-0.09	0.06	0.91	-1.43	.153	.344
GGC[a]: f vs. b × Subjective competence	0.03	0.07	1.03	0.42	.676	.838
GGC[a]: m vs. b × Subjective competence	0.19	0.06	1.21	3.04	.002	.009
GGC[a]: f vs. b × Conscientiousness	0.06	0.07	1.06	0.88	.379	.568
GGC[a]: m vs. b × Conscientiousness	-0.03	0.07	0.97	-0.39	.698	.838
GGC[a]: f vs. b × Extraversion	0.07	0.07	1.07	1.04	.299	.501
GGC[a]: m vs. b × Extraversion	0.08	0.06	1.08	1.18	.239	.478
GGC[a]: f vs. b × Risk-taking tendency	-0.18	0.07	0.84	-2.58	.010	.036
GGC[a]: m vs. b × Risk-taking tendency	-0.10	0.06	0.91	-1.54	.125	.321

Note. Hosmer and Lemeshow R^2 = .02, Cox and Snell R^2 = .02, Nagelkerke R^2 = .03. $\chi^2(17) = 169.29$, $p < .001$. $N = 10377$. BH-adjusted p = Benjamini-Hochberg adjusted p-value. GGC = Group gender composition. All predictors were centered on their group means and z-standardized.
[a]The factor group gender composition was dummy-coded with mixed-gender groups as the baseline (b) category, compared to either female (f) or male groups (m).

Follow-up questions. For exploratory purposes, we examined whether perceived leadership initiative, motivation during the group task, and group identification differed between the group gender compositions and studies. Therefore, we conducted a 2 (Study: 1

vs. 2) by 3 (group gender composition: mixed-gender vs. male vs. female groups) analysis of variance (ANOVA) for each of the three measures of the follow-up questionnaire. We report the results in the following.

The ANOVA for perceived leadership initiative did not reveal a significant effect for Study, $F(1, 766) = 0.33$, $p = .566$, $\eta_p^2 = .00$, indicating that perceived leadership initiative did not differ significantly between Study 1 ($M = 4.15$, $SD = 0.62$) and Study 2 ($M = 4.12$, $SD = 0.64$). Moreover, we did not find a significant effect for group gender composition, $F(2, 766) = 2.05$, $p = .130$, $\eta_p^2 = .01$, indicating that perceived leadership initiative did not differ significantly between mixed-gender, male, and female groups ($M = 4.11$, $SD = .61$, vs. $M = 4.09$, $SD = 0.62$, vs. $M = 4.19$, $SD = .66$). The interaction between study and group gender composition was also not significant, $F(2, 766) = 2.02$, $p = .133$, $\eta_p^2 = .01$. Therefore, the perceived amount of leadership initiative a participant showed did not differ between the two studies or between the different group gender compositions.

The ANOVA for the motivation during the group task revealed a significant effect for study, $F(1, 764) = 40.92$, $p < .001$, $\eta_p^2 = .05$, indicating that the motivation in the group was lower in Study 1 ($M = 4.62$, $SD = 0.05$), compared to Study 2 ($M = 5.05$, $SD = 0.05$). Furthermore, the analysis showed a significant effect for group gender composition, $F(2, 764) = 13.21$, $p < .001$, $\eta_p^2 = .03$. Post-hoc comparisons revealed that the motivation during the group task was higher in mixed-gender groups ($M = 4.97$, $SD = 0.93$) than in female groups ($M = 4.58$, $SD = 1.05$), $t(495.82) = 4.42$, BH-adjusted $p < .001$, $d = 0.39$, and higher in male groups ($M = 4.94$, $SD = 0.95$) than in female groups, $t(502.92) = 4.11$, BH-adjusted $p < .001$, $d = 0.36$. The motivation during the group task did not differ between mixed-gender and male groups, $t(515.99) = 0.31$, BH-adjusted $p = .757$, $d = 0.03$. The interaction between study and group gender composition was not significant, $F(2, 764) = 4.13$, $p = .103$, $\eta_p^2 = .01$. Compared to Study 1, participants were more motivated in Study 2, where they were able to

engage in a group discussion while working together. Moreover, mixed-gender and male groups were more motivated than female groups.

Additionally, an ANOVA for group identification revealed a significant main effect for study, $F(1, 764) = 40.92$, $p < .001$, $\eta_p^2 = .05$, indicating that the group identification was lower in Study 1 ($M = 3.88$, $SD = 1.23$), compared to Study 2 ($M = 4.38$, $SD = 1.20$). There was no significant main effect for group gender composition, $F(2, 764) = 2.42$, $p = .090$, $\eta_p^2 = .01$, indicating that the mean group identification did not differ between mixed-gender, male, and female groups, ($M = 4.25$, $SD = 1.20$ vs. $M = 4.07$, $SD = 1.27$ vs. $M = 4.04$, $SD = 1.26$). The interaction between study and group gender was also not significant, $F(2, 764) = 0.63$, $p = .553$, $\eta_p^2 = .00$. In sum, compared to Study 1, the group identification was higher in Study 2 where the groups engaged in a discussion while working together.

For exploratory purposes, we examined whether there were associations between our follow-up measures and leadership initiative and success of leadership, expressed as the earlier described success ratio. Table 14 shows averaged correlations across both studies. Results reveal that there is a significant correlation between leadership initiative and the motivation during the group task, averaged $r = .22$, that was significant in both studies, both $ps < .001$. Therefore, the more motivated a person was, the more leadership initiative she showed. The other correlations did not reach conventional levels of significance.

Table 14

Averaged correlations between follow-up measures and dependent variables for Study 1 and 2

Variable	1	2	3	4
1. Leadership initiative				
2. Success ratio	.16**/			
3. Perceived leadership	.08	.09		
4. Motivation during group task	.22**/**	.14	.14	
5. Group identification	.08	.13	.15	.54

Note. For success ratio, the sample size is $N = 332$ in Study 1 and $N = 299$ in Study 2. In Study 2, sample sizes ranged from $N = 378$ (for motivation during group task and group identification) to $N = 380$ (for perceived leadership).
** $p < .001$. * $p < .01$. Significance Study 1/Significance Study 2.

Discussion

The present research aimed at predicting leadership initiative and its success from individual traits, motives, and characteristics. Thereby, we addressed several research questions. First, we introduced a new view on the phenomenon emergence of leadership by differentiating two components, namely leadership initiative, on the one hand, and a leader's success in attracting followers, on the other hand. On these grounds, we investigated whether inter-individual differences might influence both components differentially. Second, we tested whether a broad range of previously identified predictors for emergence of leadership can predict behaviorally observed emergence of leadership instead of perceived/reported leadership emergence. By predicting emergence of leadership with all our predictors simultaneously, we investigated whether some of them might, for example, be dispensable because of redundancies. Third, we investigated the effects of predictors that were largely neglected by previous research, like testosterone and the implicit power motive, on leadership

initiative and its success. Fourth and finally, we explored whether group gender composition might moderate the effects of our predictors on leadership initiative and its success.

Summary of Results

Leadership initiative. Across our two studies, emergence of leadership was significantly predicted by general intelligence, subjective competence, conscientiousness, and extraversion. Therefore, the more intelligent, the more subjective competent in solving the respective task, the more conscientious, and the more extraverted a person was compared to the other members of her group, the more leadership initiative she showed. The effect of conscientiousness was moderated by study, in that it was significantly positive in the first study, whereas it was significant and even slightly negative in the second study. Therefore, some caution is warranted when interpreting this effect. Additionally, waist-to-hip-ratio was not a significant predictor in general, but its effect was significantly moderated by study, too. The effect of waist-to-hip-ratio was insignificant and positive in Study 1 but was a descriptively negative effect in Study 2. Therefore, we regard the evidence concerning this predictor as inconclusive. Investigating possible moderation effects of group gender composition, we found that subjective competence significantly predicted leadership initiative in all-male and (to a lesser extent) all-female groups, but not significantly so in mixed-gender groups. In total, we did not find any significant effects for the remaining predictors on leadership initiative.

Success of leadership initiative. Success of leadership initiative was significantly predicted by three variables, namely by general intelligence as the only significant predictor from the predictor model for leadership initiative, and then also by the accuracy of the proposed answers and the number of times a person has shown leadership initiative. Therefore, the more intelligent a person was, or the more accurate a person's suggested answers were compared to her group, the more likely it was that her proposed answer was

accepted. Moreover, the more often a person shows leadership initiative, the more likely it was that her proposed answer was accepted. Possibly, the other group members assumed that a person proposing more often an answer had good reasons to do so. Because we cannot be sure about the direction of causality, we refrain from interpreting leadership initiative as a "true" predictor for success of leadership initiative (it is, of course, a predictor in the statistical sense). We did not find any other significant effects of the investigated predictors on success of leadership initiative. We also did not find moderating effects of group gender composition on the relationships between the investigated predictors and success of leadership initiative. In other words, the above-mentioned predictors seem to be related to leadership initiative regardless of a group's gender composition.

Overall leadership. In explorative analyses, we investigated the effects of the predictors on overall leadership, operationalized as whether a person both showed leadership initiative and had success with it. Overall leadership was predicted by general intelligence, subjective competence, and extraversion. Evidence for these three predictors was rather clear. Across our two studies, we did not find a general effect of conscientiousness on overall leadership, but this effect was moderated by study, in that it had a significant positive effect in Study 1, whereas it had a significant negative effect in Study 2. Therefore, we interpret the effect of conscientiousness with caution. For overall leadership, we found that subjective competence is a significant predictor in all different gender composition groups, whereby it had the strongest effect in male groups, followed by female and mixed-gender groups. The effect of risk-taking tendency differed between mixed-gender and female groups, specifically, it was a significant predictor in mixed-gender groups, but not in female or male groups. We did not find any other significant effects predicting overall leadership.

In summary, we found evidence for the effect of general intelligence, subjective competence, and extraversion on leadership initiative. General intelligence was the only

predictor from our predictor model that had a significant effect in predicting success of leadership initiative. Overall leadership was predicted by general intelligence, subjective competence, and extraversion. General intelligence was the only predictor that had significant effects on all dependent measures of leadership emergence. Moreover, most of the significant effects were independent of the study they originated from.

Discussion of Results

In the following, we will discuss our results according to our four research goals.

First goal: Differentiating emergence of leadership into leadership initiative and its success. Our first goal was to differentially investigate the effects of inter-individual differences on the two introduced components of leadership emergence, leadership initiative, and its success. In this way, we aimed at gaining a deeper understanding of the phenomenon emergence of leadership. In the following, we will focus mainly on discussing the effects we found that are relevant for the understanding of leadership initiative and its success.

Significant predictors for leadership initiative. Previous research identified intelligence and extraversion as two important and consistently identified predictors for leadership emergence (Judge et al., 2002, 2004a). In line with our predictions, these two factors were predictive of leadership initiative in our studies, as well. The effect of these two predictors might, therefore, be considered as robust for behavioral measurements of leadership. Our paradigm provided evidence of influence on leadership emergence for both factors that is, especially for Study 1, independent of communication, like in mere group discussions.

The effect of intelligence on leadership initiative appears reasonable because intelligent individuals have the cognitive capacities to excel in problem-solving tasks. In the present studies, more intelligent persons used their ability not only to solve numerical estimation and wilderness survival tasks but also to explain their proposed answers

convincingly in order to influence the other group members. Therefore, general intelligence was beneficial in solving these tasks and being able to propose and explain an answer before another group member could seize that opportunity. General intelligence does not only play a crucial role in the present study but is important for many typical leaders' tasks (Judge et al., 2004a). Additionally, more intelligent persons might tend to show leadership behavior that is associated with initiating structure. As these leadership behaviors include initiating activities in the group and structuring how the work will be done (Bass & Bass, 2008), they match our operationalization of leadership initiative in the present research well.

Our finding that extraversion predicts leadership initiative is also plausible (and was predicted) because social and dominant people are more likely to assert themselves in a group (Judge et al., 2002). In our studies, extraversion might have been an important driving factor for showing leadership initiative. The tendency to be extraverted, and therefore, having the desire for interacting with others (Taggar et al., 1999), might have been a motivating and facilitating factor in proposing an answer in front of the group. Extraverts usually possess social confidence (Brunell et al., 2008) and, therefore, find more easily the courage to speak in front of a group. Possibly, they are also better at dealing with the fear that they might not succeed in convincing the other group members to follow them.

Among the newly investigated predictors for leadership emergence, we found an effect of subjective competence on leadership initiative. Assessing the subjective competence of an individual was largely neglected in previous research, although it is highly plausible that an individual's own assessment of her competence in solving a given task might make her show leadership initiative. Compared to extraversion and general intelligence, subjective task competence had no less of an impact on leadership initiative. Compared to general intelligence, an objective measure of the cognitive ability of a person, subjective competence reflects a person's own assessment and therefore includes the subjective task-related

confidence in solving a specific task. In other words, the competence a person perceives of herself to solve a specific kind of task is different from her actual ability to solve logical problems and might additionally motivate a person to engage in leadership behaviors requiring to solve these tasks.

The effect of conscientiousness on leadership initiative should be interpreted with caution. As one of the big five traits of personality, conscientiousness is considered as an established predictor of leadership emergence (Judge et al., 2002). In our studies, however, the effect of conscientiousness on leadership initiative was positive in Study 1 and negative in Study 2. This discrepancy in the direction of the effects might be traceable to the different group tasks in the studies. It is possible that conscientious persons hesitated more to propose an answer in the survival tasks in Study 2 because it was harder to demonstrate the correctness of a proposed solution. Compared to the estimation tasks, the wilderness survival tasks provide a scenario allowing for more creative and different possible solutions. Therefore, participants might not come up with correct solutions as straightforward as for the estimation tasks. Less conscientious persons might worry less about demonstrability and, therefore, be faster to suggest an answer they consider plausible. This is in line with how conscientiousness is usually defined: It includes striving for performance and a sense of duty, as well as the sense of responsibility. This sense of responsibility might have hindered conscientious individuals from suggesting a solution and demonstrable arguments in Study 2. These considerations suggest that conscientiousness might have a more complex relationship with leadership initiative than previously assumed. Conscientiousness might be a positive predictor for leadership initiative in easy tasks with high demonstrability, whereas it might be a negative predictor for more complex tasks with low demonstrability (e.g., tasks asking for subjective judgments). This is an interesting hypothesis that deserves further testing.

Significant predictors for success of leadership initiative. Success of leadership initiative was predicted by general intelligence. Participants exhibited rational behavior by following a more intelligent person. In a similar way, it was more likely that participants followed group members whose answers were, on average, more accurate, and who showed leadership initiative more often. Although we found an association between the odds of successful leadership initiative and the number of times a person has shown leadership initiative, we refrain from further interpreting it. It remains unclear whether the number of times a participant showed leadership initiative had a direct causal influence on its success rate, or whether, in terms of reverse causality, experiencing success increased the likelihood of trying to lead the group again.

Intelligence might be a predictor for success of leadership initiative because, as we already proposed, more intelligent individuals might have been more persuasive in explaining why their proposed answers might be correct and might have come up with more elaborate solutions. It is important to note that the effect of intelligence was still significant after the effect of accuracy was accounted for. Therefore, the effect of intelligence cannot only be explained by more intelligent persons' accurate answers. It represents a reasonable strategy to follow an intelligent individual. Intelligence is associated with problem-solving capacities, which are helping the group members to attain their goals (Bastardoz & Van Vugt, 2019; Judge et al., 2004a). Intelligence might serve as a predictor for success of leadership initiative in the present research because it becomes apparent in explaining why the proposed answer might be correct. Intelligence was a significant predictor besides of the accuracy of the proposed answers. Therefore, it was crucial that a person suggested an accurate answer and explained it convincingly.

Except for general intelligence, no other inter-individual differences reached significance in predicting success of leadership initiative. In the present studies, we observed

a constraint of variance in success of leadership. Most of the proposed answers were accepted as the group's decision. This constraint made it hard to detect effects for any of our predictors. In Study 2, even more of the proposed answers were accepted than in Study 1. Participants seemed to come to an agreement even more often when they were able to discuss the proposed answer. Therefore, before dismissing the effect of the other predictors on success of leadership initiative, these findings should be replicated using a different task, allowing for more variability in the acceptance of the proposed answers. In the present research, participants publicly showed whether they agreed with a proposed answer. If participants could make their decisions privately, they might reject the proposed answers more often.

The results of our studies give us first empirical evidence that the differentiation of emergence of leadership in the two proposed concepts, leadership initiative, and its success, is reasonable. Considering that we predicted leadership initiative and its success using the same predictor sets with only very small variations, our results showed that success of leadership initiative was nonetheless predicted by different predictors than leadership initiative. Leadership initiative was predicted by more inter-individual differences than its success. It is important to note that the two criterion variables had different variances and, therefore, we have to interpret these diverging findings cautiously. However, a variable that is predicting whether a person shows leadership initiative did not necessarily predict whether that person was successful with her leadership attempt in both our studies. To further explore that consideration, we looked at descriptive effects of our predictors on success of leadership initiative. By applying our predictor model to success of leadership initiative, we find positive descriptive effects for self-monitoring, risk-taking tendency, and testosterone, and a descriptively negative effect of conscientiousness. Because the effects of these predictors did not reach significance, we only interpret them with caution. Nonetheless, there might be more

predictors for success of leadership initiative, that were not predictive for leadership initiative. These and our significant effects should be replicated by future research using a paradigm that allows for more variance in the criterion.

To conclude, our results indicate that predictors have a differential effect on leadership initiative and its success and therefore, suggest that this differentiation of emergence of leadership is not only theoretically relevant but also on an empirical level. However, considering the constraint in variance for success of leadership initiative, other paradigms need to be tested that allow for more variance in the criterion to replicate and possibly extend these findings.

Significant predictors for overall leadership. Although it is important to conceptually differentiate between leadership initiative and its success, in practice, the combination of both variables is important because only persons who are successful in their leadership attempts will eventually lead. Our research revealed that showing overall leadership is predicted by general intelligence, subjective competence, and extraversion. The effect of conscientiousness failed to reach conventional levels of significance and was moderated by study, indicating a less robust effect (which is why we think that the effect should be replicated first before it can be further interpreted). We found the same predictors for overall leadership as for leadership initiative, including the effect of general intelligence that is also predictive of success of leadership initiative. Because we do not find evidence for additional significant predictors, the results confirm the effects of these predictors on our criteria.

To conclude, we found evidence that several predictors are predictive for leadership initiative, whereas success of leadership initiative is predicted only by general intelligence as an inter-individual difference. Although we tested a broad range of predictors, our results showed that many predictors did not reach significant effects for emergence of leadership. This might be due to several reasons that we discuss below.

Second goal: Behavioral measurements of leadership initiative and its success. In the present research, we used a broad range of inter-individual differences to predict the emergence of leadership, and we assessed emergence of leadership behaviorally. We did not find evidence for multiple predictors which had previously been identified as having an influence on leadership emergence (e.g., self-monitoring, self-esteem, dominance, or narcissism). In the following, we will discuss why we might not have been able to show significant effects of many of these predictors on leadership emergence.

We see, at least, three possible explanations for why many of the investigated predictors did not yield significant effects. The first one is that previous studies showing significant effects of predictors used subjective ratings from superiors and group members as a measure of emergence of leadership. These perceptual measurements might be an expression of raters' implicit leadership theories rather than actual leadership emergence. A person might have matched the other group members' internalized prototype of a leader and, therefore, might have been considered as a leader during the group work, independent of her behavior. Therefore, previously significant predictors might rather show which traits match a followers' implicit leader prototypes rather than predicting which person has actually emerged as a leader in a group. Consider the following example: A group is working on a task and thereby engages in a group discussion. A dominant member talks the most but is actually only saying empty phrases without any benefit towards helping the group accomplishing the task. In subsequent perceptual leadership ratings, this person might receive high scores because the group members hold prototypes that a leader dominates discussions and talks a lot. Because we used a behavioral measurement of emergence of leadership that does not underlie perceptual influences, we might have found a reduced number of significant predictors that relate to actual leadership emergence instead of leader prototypes.

A second explanation for why we did not find effects for many previously identified inter-individual differences might be that we used an especially broad range of predictors for predicting leadership initiative and its success. Previous research mostly focused on assessing a few predictors, and studies typically used different sets of predictors. Hence, there might be overlapping and therefore, redundant predictors in the literature. In order to examine which predictors have a unique predictive value, we simultaneously used them in our broad predictor model. As extraversion was a significant predictor for leadership initiative, we will use it in the following as an example for the identification of redundancies: Belonging to the big five model of personality, previous research suggests that extraversion is a broad personality trait that might cover several different other, more narrow personality facets (Judge et al., 2002). Extraversion is considered to include the tendency to be dominant (Judge et al., 2002), while dominance is often used as a separate predictor for emergence of leadership, without measuring extraversion simultaneously (e.g., Foti & Hauenstein, 2007). In our studies, we find that extraversion is strongly correlated with dominance across both studies (averaged $r = .51$, both $ps < .001$). Moreover, within our studies, extraversion is substantially correlated with narcissism (averaged $r = .46$, both $ps < .001$). Although there are studies reporting that narcissism predicts leadership emergence beyond the effect of extraversion (e.g., Brunell et al., 2008), a recent meta-analysis shows that the relationship between narcissism and leadership emergence can be explained by the overlap of narcissism with extraversion (Grijalva et al., 2015). Our results are in line with that conclusion in that narcissism is not a significant predictor for leadership initiative when we simultaneously include extraversion as a predictor. To sum up, the results of our studies suggest that extraversion might be sufficient to cover effects from related personality constructs, that is, dominance, and narcissism. Because of similar reasons, other predictors might not have reached significance. Future research is needed to replicate these findings.

A third explanation for why we did not find effects for many predictors might be due to the design of our paradigm. Within the present research, we aimed at measuring the two components leadership initiative and its success as unambiguously and as independent from perceptional biases as possible. For that purpose, we decided to use a highly controlled paradigm with a thoroughly pre-structured task in the laboratory. Within this design, we limited leadership behavior to merely proposing an answer to a choice task, and consequently, that was the only behavior we measured as leadership behavior. That means that even if participants showed other forms of leadership behavior, we did not measure them. As a hypothetical example, consider a person proposed an answer in Study 2. After she finished her explanation, another group member praised her for her performance. In this case, this might have been leadership behavior driven by agreeableness, that is not included in our measurement of leadership behavior. Therefore, some predictors possibly were not identified as significant predictors, that might show effects in other types of paradigms that involve other forms of leader behavior. Additionally, during the group phase, we restricted participants' amount of communication prior to suggesting an answer in Study 1, and an additional pre-structured discussion in Study 2. If we would apply our measurement of leadership initiative and its success to that scenario, leadership behaviors based on communication would also be missed. We discuss the advantages and disadvantages of the present paradigm in a following section regarding limitations and directions for future research.

In sum, future research might provide further insight into which predictors are not predictive of emergence of leadership and investigate whether this is due to controlled research paradigms or the behavioral assessment of emergence of leadership. At this point, it is important to note that, at the very least, there is one category of predictors where the

missing effects can hardly be explained by our measurement of emergence of leadership – namely the physical features. We will discuss this in the following.

We did not find robust evidence for the effect of physical features on leadership emergence in both of our studies. An advantage of our study is that we base our findings on very reliable data. In the present research, we measured physical features very precisely by using a 3D-body scanner for the assessment of the waist-to-hip-ratio. In previous research, waist-to-hip-ratio was measured manually by research assistants (e.g., Campbell et al., 2002), which is much more error-prone than automatic measurements. Previous findings showing associations between physical features and leadership emergence (e.g., Campbell et al., 2002; Judge & Cable, 2004) might be the result of perceptual measurements of emergent leaders. As argued by Blaker et al. (2013), this can be a manifestation of the so-called "mismatch effect": Persons with desirable physical features linked to evolutionary fitness might be automatically and unconsciously perceived as leaders. Stated differently, previous research might have shown that the effect of physical features on emergence of leadership reflected persons' implicit leadership theories on how they imagine a prototypical leader instead of measuring actual leadership behavior. Therefore, our findings underline that behavioral measurement is crucial.

Third goal: Newly identified predictors and approaches in measuring predictors. Within the present research, we aimed at investigating whether newly identified predictors and approaches in measuring predictors have an effect on leadership initiative and its success. New predictors for emergence of leadership included testosterone, risk-taking tendency, and subjective competence. Additionally, we assessed the power motive implicitly, via the Picture Story Exercise, and explicitly, via the recently developed dominance, prestige, and leadership scales (Suessenbach et al., 2018). In summary, subjective task competence was a significant predictor of leadership initiative in both studies. Risk-taking tendency was a

significant predictor for overall leadership in mixed-gender groups. We did not find significant effects of the other predictors on our criteria. In the present research, we measured leadership initiative and its success very unambiguously, and these predictors might not have reached significance because of the already discussed reasons. Future research might replicate these effects.

Fourth goal: Group gender composition. Our fourth goal was to exploratively investigate the potential moderating effects of group gender composition on the relation between our predictor set, on the one hand, and the emergence of leadership, on the other hand. Thereby, we investigated whether group gender composition differentially influenced leadership initiative and its success. Moreover, we aimed at investigating whether the effects of our predictors were robust. Predicting leadership initiative, we found that subjective competence was a significant predictor in male and female groups, with the effect being stronger in male groups. In contrast, subjective competence was not a significant predictor in mixed-gender groups. We did not find any further significant interactions between predictors and success of leadership initiative. For overall leadership, subjective competence was a significant predictor in all three gender composition groups. The effect was strongest in male groups, followed by female groups, and was weakest in mixed-gender groups. Moreover, risk-taking tendency was a significant predictor in mixed-gender groups, whereas it was not a significant predictor in same-gender groups.

Compared to the number of predictors we assessed in this study, we found relatively few interaction effects between group gender composition and the predictors on our criterion variables. Overall, we consider the effects of our predictor variables as robust. We discuss the effects that we found in the following.

Subjective competence might have had the strongest effect on leadership initiative and overall leadership in same-gender groups because participants wanted to avoid proposing

a wrong answer option. Therefore, they may not have made a leadership attempt when they did not feel subjectively competent.

In the case of mixed-gender groups, the effect of a person's risk-taking tendency on overall leadership might be due to the fact that both men and women were involved. Perhaps, in order to impress the other gender, participants prone to engage in risky behaviors might have been more motivated to make a leadership attempt that was also successful. Moreover, participants may have been more scared to fail with their leadership attempts. In same gender groups, participants may have perceived it as not risky to show leadership initiative, and therefore, risk-taking tendency was not a significant predictor for overall leadership.

Limitations and Implications for Future Research

One limitation of the present research is that we only measured a restricted amount of leadership behavior due to the design of our studies. This limitation raises three issues that we will address in the following. First, we will discuss the advantages and limitations of the present paradigm. Second, and related to that, we will discuss the extent of leadership behavior that we were able to measure within the present studies. Third, as a consequence, we will present a research idea for future studies, addressing the apparent criticism.

First, the present paradigm offered considerable advantages and limitations. Because we presented a new approach to investigating the emergence of leadership, we aimed at measuring the two proposed components leadership initiative and its success as unambiguously as possible and also behaviorally. In order to achieve that research goal, we designed the study in a way that guarantees that emergence of leadership occurs and we thoroughly pre-structured the group interactions. Thereby, we limited the range of leadership behaviors and the amount of communication our participants could show.

Simultaneously, that approach comes with the disadvantage that the group interaction was rather artificial. Although in practice, groups are usually free to decide on how they work

on tasks and discuss proposed answers, the work in the group was pre-structured, and the amount of communication was limited in the present paradigm.

Less controlled paradigms might offer potential benefits, but they would have several disadvantages that might be detrimental for our research purposes. To illustrate this, consider a hypothetical paradigm in which participants are able to freely communicate while working on the group task. On first thought, this approach is beneficial. The work in the group would be more realistic because groups are usually free to decide on how to solve a task and are able to communicate naturally and without any restrictions. However, if participants were free in communicating and engaging in a discussion, the measurement of leadership initiative and its success becomes blurred easily. If we would apply our measurement of leadership initiative and its success to that scenario, leadership behaviors based on communication would be missed. For example, participants could show leadership initiative and its success verbally, and we would not have detected this in our present leadership emergence measures. Therefore, we would risk introducing confounding factors if we allowed participants to communicate freely. Even within the controlled paradigm of the present research, there is an example of how communication interferes with the primary aims of the study. In Study 1, a few participants deliberately proposed an answer they considered as wrong and explicitly asked their group members to disagree with this answer. This behavior worked against our operationalization of leadership initiative and its success. It is likely that these kinds of transgressions occur more often in less controlled settings. Thus, we considered our studies' design as essential in order to fulfill the primary research aims.

A second limitation of the present research might be the restricted leadership behaviors we allowed for within our paradigm. Considering the limited opportunities with regard to the behaviors and actions our participants could show, it might be debated whether and to what extent we really measured leadership in our studies. At its core, and as outlined

in the theoretical introduction, leadership is defined as the exertion of influence on others in order to reach a common goal (House et al., 2004). Apparently, leadership behavior can consist of many facets. Within leadership research, initiating structure and consideration are assumed to be the two most fundamental types of leadership behaviors. In general, initiating structure refers to leader behaviors that are oriented towards the task and goal attainment (Bass & Bass, 2008; Fleishman, 1953). This dimension of leadership behavior describes the extent to which a leader initiates, organizes, and structures work in a group, and whether she insists on meeting standards and deadlines. Leaders showing initiating structure decide in detail what will be done and how it will be done. This leadership behavior is characterized by clear communication (Bass & Bass, 2008). In contrast, consideration describes the extent to which a leader expresses concern for the welfare of his group members. Considerate leadership behaviors include expressing appreciation for good work, treating subordinates as equals, and considering group members' suggestions. Considerate leaders are easy to approach, establish relationships based on mutual trust, and let group members participate in decisions (Bass & Bass, 2008). Compared to initiating structure, considerate leadership behaviors seem to be more grounded in communication (De Vries, Bakker-Pieper, & Oostenveld, 2010). In our study, leadership initiative represents leadership behaviors more associated with initiating structure. Within our paradigm, participants showed leadership by actively proposing an answer to help the group reach its goal. Moreover, instead of only stating a preference, they provided arguments in order to influence and convince the other group members of their leadership attempt. This behavior was directed towards a shared goal, that is, solving the given tasks and achieving a high performance-based bonus. Although there are several facets of this behavior, providing a strategy to solve a task is a way to show strategies for goal attainment. On the other hand, participants' opportunities to show

considerate leadership behaviors were not only much more limited but, most importantly, we did not consider them in our measurement of leadership initiative.

Third, the considerations on the controlled nature of our paradigm and the restricted measurement of leadership behavior lead us to directions for future research that might use another paradigm to measure leadership emergence. Ideally, a new paradigm increases the group interaction's realism and measures more differentiated facets of leadership behavior, for example, by considering the fundamental leadership behaviors initiating structure and consideration. At the same time, and ideally, it should still precisely measure leadership initiative and its success on a behavioral level. Within the paradigm, it is crucial that the participants' behavior is directed so leadership initiative, and its success can be measured. Within this approach, two clear coding systems for leadership behaviors are needed, namely one for initiating structure and one for consideration. Within each of these coding systems, leadership initiative and its success need to be included. For initiating structure, additional behaviors might be included in the measurement. An example of leadership initiative might be a suggestion on a plan on how to structure a given task, while its success might be measured as whether that plan is put into action or not. In the case of considerate leader behaviors, for example, leadership initiative might take the shape of a participant offering to care for someone during the group work, and its success might be operationalized as whether that offer is accepted or not. Future research using a new paradigm offers the opportunity to replicate the findings of the present research, but also might explore whether we find effects of predictors that did not reach significance in the present research.

Conclusions

The present research investigated the emergence of leadership and how we can predict its components (leadership initiative and success of leadership initiative) from individual traits, motives, and characteristics. Our findings contribute to a deeper

understanding of the emergence of leadership by showing that the two components can be predicted differentially. We showed that general intelligence, subjective competence, and extraversion robustly predicted leadership initiative, whereas its success was predicted by general intelligence and the accuracy of the proposed answers. From a broad range of previously identified predictors, we were able to confirm the robust effects of a few with a behavioral instead of a perceptual measurement of emergence of leadership, namely general intelligence and extraversion. The failure to replicate the effects of the remaining predictors that are known from the literature might be due to implicit leadership theories or due to the controlled nature of our research paradigm (or due to reasons that we have not considered so far). Clarifying these reasons is a challenge for further research. Moreover, the present research suggests that physical features do not predict leadership initiative and its success. This is noteworthy because the lack of findings cannot be attributed to the fact that the employed research paradigm offered limited opportunities for communication. Moreover, we did not find a huge body of evidence for the moderating effect of group gender composition on the relation between our predictor set and the emergence of leadership. The present research contributed to research on leadership emergence by differentiating it into leadership initiative and its success and offers avenues for future research to deepen this understanding.

References

Ackerman, R. A., Witt, E. A., Donnellan, M. B., Trzesniewski, K. H., Robins, R. W., & Kashy, D. A. (2011). What does the narcissistic personality inventory really measure? *Assessment*, *18*(1), 67–87. https://doi.org/10.1177/1073191110382845

Acton, B. P., Foti, R. J., Lord, R. G., & Gladfelter, J. A. (2019). Putting emergence back in leadership emergence: A dynamic, multilevel, process-oriented framework. *The Leadership Quarterly*, *30*(1), 145–164. https://doi.org/10.1016/j.leaqua.2018.07.002

Ahuja, A., & Van Vugt, M. (2010). *Selected: Why some people lead, why others follow, and why it matters*. Profile Books. Retrieved from https://market.android.com/details?id=book-rUol1wv_0NwC

Amos, B., & Klimoski, R. J. (2014). Courage: Making teamwork work well. *Group & Organization Management*, *39*(1), 110–128. https://doi.org/10.1177/1059601113520407

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

Antonakis, J., & Dalgas, O. (2009). Predicting elections: Child's play! *Science*, *323*(5918), 1183. https://doi.org/10.1126/science.1167748

Antonakis, J., & Eubanks, D. L. (2017). Looking leadership in the face. *Current Directions in Psychological Science*, *26*(3), 270–275. https://doi.org/10.1177/0963721417705888

Antonakis, J., Ashkanasy, N. M., & Dasborough, M. T. (2009). Does leadership need emotional intelligence? *The Leadership Quarterly*, *20*(2), 247–261. https://doi.org/10.1016/j.leaqua.2009.01.006

Archer, J. (2006). Testosterone and human aggression: An evaluation of the challenge hypothesis. *Neuroscience and Biobehavioral Reviews*, *30*(3), 319–345. https://doi.org/10.1016/j.neubiorev.2004.12.007

Atwater, L. E., Dionne, S. D., Avolio, B., Camobreco, J. F., & Lau, A. W. (1999). A longitudinal study of the leadership development process: Individual differences predicting leader effectiveness. *Human Relations*, *52*(12), 1543–1562. https://doi.org/10.1177/001872679905201203

Back, M. D., Küfner, A. C. P., Dufner, M., Gerlach, T. M., Rauthmann, J. F., & Denissen, J. J. A. (2013). Narcissistic admiration and rivalry: Disentangling the bright and dark sides of narcissism. *Journal of Personality and Social Psychology*, *105*(6), 1013–1037. https://doi.org/10.1037/a0034431

Barney, J. B., & Hansen, M. H. (1994). Trustworthiness as a source of competitive advantage. *Strategic Management Journal*, *15*, 175–190. https://doi.org/10.1002/smj.4250150912

Bass, B. M., & Bass, R. (2008). *Handbook of leadership: Theory, research, and managerial applications* (4th ed.). New York, NY: Free Press.

Bastardoz, N., & Van Vugt, M. (2019). The nature of followership: Evolutionary analysis and review. *The Leadership Quarterly*, *30*(1), 81–95. https://doi.org/10.1016/j.leaqua.2018.09.004

Ben-Ner, A., & Halldorsson, F. (2010). Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology*, *31*(1), 64–79. https://doi.org/10.1016/j.joep.2009.10.001

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series*

B (Methodological), *57*(1), 289–300. Retrieved from http://www.jstor.org/stable/2346101

Berggren, N., Jordahl, H., & Poutvaara, P. (2010). The looks of a winner: Beauty and electoral success. *Journal of Public Economics*, *94*(1-2), 8–15. https://doi.org/10.1016/j.jpubeco.2009.11.002

Berkowitz, L. E. (1979). *Advances in experimental social psychology* (Vol. 12). New York: Academic Press.

Beyan, C., Carissimi, N., Capozzi, F., Vascon, S., Bustreo, M., Pierro, A., … Murino, V. (2016). Detecting emergent leader in a meeting environment using nonverbal visual features only. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction* (pp. 317–324). New York, NY: ICIM'16; ACM. https://doi.org/10.1145/2993148.2993175

Beyan, C., Katsageorgiou, V.-M., & Murino, V. (2017). Moving as a leader: Detecting emergent leadership in small groups using body pose. In *Proceedings of the 2017 ACM on Multimedia Conference* (pp. 1425–1433). New York, NY: ACM. https://doi.org/10.1145/3123266.3123404

Blaker, N. M., Rompa, I., Dessing, I. H., Vriend, A. F., Herschberg, C., & Van Vugt, M. (2013). The height leadership advantage in men and women: Testing evolutionary psychology predictions about the perceptions of tall leaders. *Group Processes & Intergroup Relations*, *16*(1), 17–27. https://doi.org/10.1177/1368430212437211

Boyer, T. W. (2006). The development of risk-taking: A multi-perspective review. *Developmental Review*, *26*(3), 291–345. https://doi.org/10.1016/j.dr.2006.05.002

Brunell, A. B., Gentry, W. A., Campbell, W. K., Hoffman, B. J., Kuhnert, K. W., & DeMarree, K. G. (2008). Leader emergence: The case of the narcissistic leader.

Personality and Social Psychology Bulletin, *34*(12), 1663–1676. https://doi.org/10.1177/0146167208324101

Campbell, L., Simpson, J. A., Stewart, M., & Manning, J. G. (2002). The formation of status hierarchies in leaderless groups : The role of male waist-to-hip ratio. *Human Nature*, *13*(3), 345–362. https://doi.org/10.1007/s12110-002-1019-0

Campbell, L., Simpson, J. A., Stewart, M., & Manning, J. G. (2003). Putting personality in social context: extraversion, emergent leadership, and the availability of rewards. *Personality & Social Psychology Bulletin*, *29*(12), 1547–1559. https://doi.org/10.1177/0146167203256920

Cherulnik, P., Turns, L. C., & Wilderman, S. K. (1990). Physical appearance and leadership: Exploring the role of appearance based attribution in leader emergence. *Journal of Applied Social Psychology*, *20*(18), 1530–1539. https://doi.org/10.1111/j.1559-1816.1990.tb01491.x

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Colbert, A. E., Judge, T. A., Choi, D., & Wang, G. (2012). Assessing the trait theory of leadership using self and observer ratings of personality: The mediating role of contributions to group success. *The Leadership Quarterly*, *23*(4), 670–685. https://doi.org/10.1016/j.leaqua.2012.03.004

Conger, J. A., & Kanungo, R. N. (1998). *Charismatic leadership in organizations*. Thousand Oaks, CA: Sage.

Côté, S., Lopes, P. N., Salovey, P., & Miners, C. T. H. (2010). Emotional intelligence and leadership emergence in small groups. *The Leadership Quarterly*, *21*(3), 496–508. https://doi.org/10.1016/j.leaqua.2010.03.012

Crockett, W. H. (1955). Emergent leadership in small, decision-making groups. *The Journal of Abnormal and Social Psychology*, *51*, 378–383. https://doi.org/10.1037/h0046109

Cronshaw, S. F., & Ellis, R. J. (1991). A process investigation of self-monitoring and leader emergence. *Small Group Research*, *22*(4), 403–420. https://doi.org/10.1177/1046496491224001

Cronshaw, S. F., & Lord, R. G. (1987). Effects of categorization, attribution, and encoding processes on leadership perceptions. *Journal of Applied Psychology*, *72*(1), 97–106. https://doi.org/10.1037/0021-9010.72.1.97

De Vries, R. E., Bakker-Pieper, A., & Oostenveld, W. (2010). Leadership = communication? The relations of leaders' communication styles with leadership styles, knowledge sharing and leadership outcomes. *Journal of Business and Psychology*, *25*(3), 367–380. https://doi.org/10.1007/s10869-009-9140-2

Dekkers, T. J., Rentergem, J. A. A. van, Meijer, B., Popma, A., Wagemaker, E., & Huizenga, H. M. (2019). A meta-analytical evaluation of the dual-hormone hypothesis: Does cortisol moderate the relationship between testosterone and status, dominance, risk taking, aggression, and psychopathy? *Neuroscience and Biobehavioral Reviews*, *96*, 250–271. https://doi.org/10.1016/j.neubiorev.2018.12.004

Diz, A. P., Carvajal-Rodríguez, A., & Skibinski, D. O. F. (2011). Multiple hypothesis testing in proteomics: a strategy for experimental work. *Molecular & Cellular Proteomics: MCP*, *10*(3), M110.004374. https://doi.org/10.1074/mcp.M110.004374

Doosje, B., Ellemers, N., & Spears, R. (1995). Perceived intragroup variability as a function of group status and identification. *Journal of Experimental Social Psychology*, *31*(5), 410–436. https://doi.org/10.1006/jesp.1995.1018

Eby, L. T., Cader, J., & Noble, C. L. (2003). Why do high self-monitors emerge as leaders in small groups? A comparative analysis of the behaviors of high versus low self-

monitors. *Journal of Applied Social Psychology*, *33*(7), 1457–1479. https://doi.org/10.1111/j.1559-1816.2003.tb01958.x

Efrain, M. G., & Patterson, E. W. J. (1974). Voters vote beautiful: The effect of physical appearance on a national election. *Canadian Journal of Behavioural Science.*, *6*(4), 352–356. https://doi.org/10.1037/h0081881

Ellis, R. J. (1988). Self-Monitoring and leadership emergence in groups. *Journal of Composite Materials*, *33*(10), 928–940. https://doi.org/0803973233

Emery, C., Calvard, T. S., & Pierce, M. E. (2013). Leadership as an emergent group process: A social network study of personality and leadership. *Group Processes & Intergroup Relations*, *16*(1), 28–45. https://doi.org/10.1177/1368430212461835

Ensari, N., Riggio, R. E., Christian, J., & Carslaw, G. (2011). Who emerges as a leader? Meta-analyses of individual differences as predictors of leadership emergence. *Personality and Individual Differences*, *51*(4), 532–536. https://doi.org/10.1016/j.paid.2011.05.017

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Ferring, D., & Filipp, S. H. (1996). Messung des Selbstwertgefühls: Befunde zu Reliabilität, Validität und Stabilität der Rosenberg-Skala. *Diagnostica*, *42*(3), 284–292.

Fischer, J., & Zinner, D. (2011). Communication and cognition in primate group movement. *International Journal of Primatology*, *32*(6), 1279–1295. https://doi.org/10.1007/s10764-011-9542-7

Fleishman, E. A. (1953). Leadership climate, human relations training, and supervisory behavior. *Personnel Psychology*, *6*(2), 205–222. https://doi.org/ https://doi.org/10.1111/j.1744-6570.1953.tb01040.x

Foster, J. D., Shenesey, J. W., & Goff, J. S. (2009). Why do narcissists take more risks? Testing the roles of perceived risks and benefits of risky behaviors. *Personality and Individual Differences*, *47*(8), 885–889. https://doi.org/10.1016/j.paid.2009.07.008

Foti, R. J., & Hauenstein, N. M. A. (2007). Pattern and variable approaches in leadership emergence and effectiveness. *Journal of Applied Psychology*, *92*(2), 347–355. https://doi.org/10.1037/0021-9010.92.2.347

Gershenoff, A. B., & Foti, R. J. (2003). Leader emergence and gender roles in all-female groups: A contextual examination. *Small Group Research*, *34*(2), 170–196. https://doi.org/10.1177/1046496402250429

Gomulya, D., Wong, E. M., Ormiston, M. E., & Boeker, W. (2017). The role of facial appearance on CEO selection after firm misconduct. *The Journal of Applied Psychology*, *102*(4), 617–635. https://doi.org/10.1037/apl0000172

Grijalva, E., Harms, P. D., Newman, D. A., Gaddis, B. H., & Fraley, R. C. (2015). Narcissism and leadership: A meta-analytic review of linear and nonlinear relationships. *Personnel Psychology*, *68*(1), 1–47. https://doi.org/10.1111/peps.12072

Gruber, F. M., Veidt, C., & Ortner, T. M. (2018). Women who emerge as leaders in temporarily assigned work groups: Attractive and socially competent but not babyfaced or naïve? *Frontiers in Psychology*, *9*, 2553. https://doi.org/10.3389/fpsyg.2018.02553

Guastello, S. J. (2007). Non-linear dynamics and leadership emergence. *The Leadership Quarterly*, *18*(4), 357–369. https://doi.org/10.1016/j.leaqua.2007.04.005

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, *78*(5), 837–852. https://doi.org/10.1037/0022-3514.78.5.837

Hogan, R., Curphy, G. J., & Hogan, J. (1994). What we know about leadership: Effectiveness and personality. *The American Psychologist*, *49*(6), 493–504. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8042818

Hong, Y., Catano, V. M., & Liao, H. (2011). Leader emergence: The role of emotional intelligence and motivation to lead. *Leadership & Organization Development Journal*, *32*(4), 320–343. https://doi.org/10.1108/01437731111134625

House, R. J. (1971). A path goal theory of leader effectiveness. *Administrative Science Quarterly*, *16*(3), 321–339. https://doi.org/10.2307/2391905

House, R. J. (1996). Path-goal theory of leadership: Lessons, legacy, and a reformulated theory. *The Leadership Quarterly*, *7*(3), 323–352. https://doi.org/10.1016/s1048-9843(96)90024-7

House, R. J., & Aditya, R. N. (1997). The social scientific study of leadership: Quo vadis? *Journal of Management*, *23*(3), 409–473. https://doi.org/10.1177/014920639702300306

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. SAGE Publications. Retrieved from https://market.android.com/details?id=book-4MByAwAAQBAJ

Jackson, D. N. (1967). *Manual for the personality research form*. Goshen, New York: Research Psychologists Press.

Jackson, D. N. (1974). *Personality research form manual*. Goshen, New York: Research Psychologists Press.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp.

114–158). New York, NY: Guilford Press. https://doi.org/10.1016/S0191-8869(97)81000-8

Johnson, J., Wilke, A., & Weber, E. U. (2004). Beyond a trait view of risk taking: A domain-specific scale measuring risk perceptions, expected benefits, and perceived-risk attitudes in German-speaking populations. *Polish Psychological Bulletin*, *35*, 153–172. Retrieved from https://papers.ssrn.com/abstract=1301128

Judge, T. A., & Cable, D. M. (2004). The effect of physical height on workplace success and income: Preliminary test of a theoretical model. *The Journal of Applied Psychology*, *89*(3), 428–441. https://doi.org/10.1037/0021-9010.89.3.428

Judge, T. A., & Piccolo, R. F. (2004). Transformational and transactional leadership: a meta-analytic test of their relative validity. *The Journal of Applied Psychology*, *89*(5), 755–768. https://doi.org/10.1037/0021-9010.89.5.755

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, *87*(4), 765–780. https://doi.org/10.1037/0021-9010.87.4.765

Judge, T. A., Colbert, A. E., & Ilies, R. (2004a). Intelligence and leadership: A quantitative review and test of theoretical propositions. *The Journal of Applied Psychology*, *89*(3), 542–552. https://doi.org/10.1037/0021-9010.89.3.542

Judge, T. A., Hurst, C., & Simon, L. S. (2009). Does it pay to be smart, attractive, or confident (or all three)? Relationships among general mental ability, physical attractiveness, core self-evaluations, and income. *The Journal of Applied Psychology*, *94*(3), 742–755. https://doi.org/10.1037/a0015497

Judge, T. A., Piccolo, R. F., & Ilies, R. (2004b). The forgotten ones? The validity of consideration and initiating structure in leadership research. *Journal of Applied Psychology*, *89*(1), 36–51. https://doi.org/10.1037/0021-9010.89.1.36

Kickul, J., & Neuman, G. (2000). Emergent leadership behaviors: The function of personality and cognitive ability in determining teamwork performance and KSAs. *Journal of Business and Psychology*, *15*(1), 27–52. https://doi.org/10.1023/A:1007714801558

King, A. J., Douglas, C. M. S., Huchard, E., Isaac, N. J. B., & Cowlishaw, G. (2008). Dominance and affiliation mediate despotism in a social primate. *Current Biology: CB*, *18*(23), 1833–1838. https://doi.org/10.1016/j.cub.2008.10.048

Kirkpatrick, S. A., & Locke, E. A. (1991). Leadership: Do traits matter? *Academy of Management Executive*, *5*(2), 48–60. https://doi.org/10.5465/ame.1991.4274679

Klimpki, J. S. (2017). *Der jeweilige Einfluss wahrgenommener Attraktivität, Kompetenz und Vertrauenswürdigkeit auf Emergence of Leadership* (Unpublished master's thesis). Georg-Elias-Müller Institute for Psychology, Göttingen.

Koellner, M. G., & Schultheiss, O. C. (2014). Meta-analytic evidence of low convergence between implicit and explicit measures of the needs for achievement, affiliation, and power. *Frontiers in Psychology*, *5*, 1–20. https://doi.org/10.3389/fpsyg.2014.00826

Kolb, J. A. (1997). Are we still stereotyping leadership? A look at gender and other predictors of leader emergence. *Small Group Research*, *28*(3), 370–393. https://doi.org/10.1177/1046496497283003

Kordsmeyer, T. L., Hunt, J., Puts, D. A., Ostner, J., & Penke, L. (2018). The relative importance of intra- and intersexual selection on human male sexually dimorphic traits. *Evolution and Human Behavior: Official Journal of the Human Behavior and Evolution Society*, *39*(4), 424–436. https://doi.org/10.1016/j.evolhumbehav.2018.03.008

Lang, F. R., Lüdtke, O., & Asendorpf, J. B. (2001). Testgüte und psychometrische Äquivalenz der deutschen Version des Big Five Inventory (BFI) bei jungen,

mittelalten und alten Erwachsenen. *Diagnostica*, *47*(3), 111–121. https://doi.org/10.1026//0012-1924.47.3.111

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, *126*(3), 390–423. https://doi.org/10.1037//0033-2909.126.3.390

Lemoine, G. J., Aggarwal, I., & Steed, L. B. (2016). When women emerge as leaders: Effects of extraversion and gender composition in groups. *The Leadership Quarterly*, *27*(3), 470–486. https://doi.org/10.1016/j.leaqua.2015.12.008

Lennox, R. D., & Wolfe, R. N. (1984). Revision of the self-monitoring scale. *Journal of Personality and Social Psychology*, *46*(6), 1349–1364. https://doi.org/10.1037/0022-3514.46.6.1349

Liening, S. H., & Josephs, R. A. (2010). It is not just about testosterone: Physiological mediators and moderators of testosterone's behavioral effects: Mediation and moderation of testosterone. *Social and Personality Psychology Compass*, *4*(11), 982–994. https://doi.org/10.1111/j.1751-9004.2010.00316.x

Liening, S. H., Stanton, S. J., Saini, E. K., & Schultheiss, O. C. (2010). Salivary testosterone, cortisol, and progesterone: Two-week stability, interhormone correlations, and effects of time of day, menstrual cycle, and oral contraceptive use on steroid hormone levels. *Physiology & Behavior*, *99*(1), 8–16. https://doi.org/10.1016/j.physbeh.2009.10.001

Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R (2. erweiterte und überarbeitete Auflage)*. Göttingen, Germany: Hogrefe.

Linke, L., Saribay, S. A., & Kleisner, K. (2016). Perceived trustworthiness is associated with position in a corporate hierarchy. *Personality and Individual Differences*, *99*, 22–27. https://doi.org/10.1016/j.paid.2016.04.076

Little, A. C., Roberts, S. C., Jones, B. C., & Debruine, L. M. (2012). The perception of attractiveness and trustworthiness in male faces affects hypothetical voting decisions differently in wartime and peacetime scenarios. *Quarterly Journal of Experimental Psychology*, *65*(10), 2018–2032. https://doi.org/10.1080/17470218.2012.677048

Lord, R. G., De Vader, C. L., & Alliger, G. M. (1986). A meta-analysis of the relation between personality traits and leadership perceptions: An application of validity generalization procedures. *Journal of Applied Psychology*, *71*(3), 402–410. https://doi.org/10.1037/0021-9010.71.3.402

Lord, R. G., Foti, R. J., & De Vader, C. L. (1984). A test of leadership categorization theory: Internal structure, information processing, and leadership perceptions. *Organizational Behavior and Human Performance*, *34*(3), 343–378. https://doi.org/10.1016/0030-5073(84)90043-6

Lovric, N. (2018). *Der jeweilige Einfluss von wahrgenommener Attraktivität, Kompetenz und Vertrauenswürdigkeit auf Emergence of Leadership* (Unpublished master's thesis). Georg-Elias-Müller Institute for Psychology, Göttingen.

Mann, R. D. (1959). A review of the relationships between personality and performance in small groups. *Psychological Bulletin*, *56*(4), 241–270. https://doi.org/10.1037/h0021468

Marinova, S. V., Moon, H., & Kamdar, D. (2013). Getting ahead or getting along? The two-facet conceptualization of conscientiousness and leadership emergence. *Organization Science*, *24*(4), 1257–1276. https://doi.org/10.1287/orsc.1120.0781

McClelland, D. C. (1987). *Human Motivation*. CUP Archive. Retrieved from https://market.android.com/details?id=book-vic4AAAAIAAJ

Mehta, P. H., Welker, K. M., Zilioli, S., & Carré, J. M. (2015). Testosterone and cortisol jointly modulate risk-taking. *Psychoneuroendocrinology*, *56*, 88–99. https://doi.org/10.1016/j.psyneuen.2015.02.023

Moss, S. E., & Kent, R. L. (1996). Gender and gender-role categorization of emergent leaders: A critical review and comprehensive analysis. *Sex Roles*, *35*(1-2), 79–96. https://doi.org/10.1007/bf01548176

Nevicka, B., De Hoogh, A. H. B., Van Vianen, A. E. M., Beersma, B., & McIlwain, D. (2011). All I need is a stage to shine: Narcissists' leader emergence and performance. *The Leadership Quarterly*, *22*(5), 910–925. https://doi.org/10.1016/j.leaqua.2011.07.011

Offermann, L. R., & Coats, M. R. (2018). Implicit theories of leadership: Stability and change over two decades. *The Leadership Quarterly*, *29*(4), 513–522. https://doi.org/10.1016/j.leaqua.2017.12.003

Offermann, L. R., Kennedy, J. K., & Wirtz, P. W. (1994). Implicit leadership theories: Content, structure, and generalizability. *The Leadership Quarterly*, *5*(1), 43–58. https://doi.org/10.1016/1048-9843(94)90005-1

Patzer, G. L. (1985). *The physical Attractiveness Phenomena*. New York: Plenum.

Paunonen, S. V., Lönnqvist, J.-E., Verkasalo, M., Leikas, S., & Nissinen, V. (2006). Narcissism and emergent leadership in military cadets. *The Leadership Quarterly*, *17*(5), 475–486. https://doi.org/10.1016/j.leaqua.2006.06.003

Pfeiffer, W., & Jones, J. E. (Eds.). (1976). *The 1976 annual handbook for group facilitators*. La Jolla, CA: University Associates.

R Core Team. (2018). R: A language and environment for statistical computing (Version 3.5.1) [Statistic Software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Raskin, R. N., & Hall, C. S. (1979). A narcissistic personality inventory. *Psychological Reports*, *45*(2), 590. https://doi.org/10.2466/pr0.1979.45.2.590

Raskin, R. N., & Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology*, *54*(5), 890–902. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/3379585

Revelle, W. (2018). Psych: Procedures for psychological, psychometric, and personality research (Version 1.8.4) [R package]. Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Rice, R. W., Yoder, J. D., Adams, J., Priest, R. F., & Prince, H. T. (1984). Leadership ratings for male and female military cadets. *Sex Roles*, *10*(11-12), 885–901.

Ronay, R., & Carney, D. R. (2013). Testosterone's negative relationship with empathic accuracy and perceived leadership ability. *Social Psychological and Personality Science*, *4*(1), 92–99. https://doi.org/10.1177/1948550612442395

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rubin, R. S., Bartels, L. K., & Bommer, W. H. (2002). Are leaders smarter or do they just seem that way? Exploring perceived intellectual competence and leadership emergence. *Social Behavior and Personality: An International Journal*, *30*(2), 105–118. https://doi.org/10.2224/sbp.2002.30.2.105

Rueb, J., Erskine, H., & Foti, R. (2008). Intelligence, dominance, masculinity, and self-monitoring: Predicting leadership emergence in a military setting. *Military Psychology: The Official Journal of the Division of Military Psychology, American Psychological Association*, *20*(4), 237–252. https://doi.org/10.1080/08995600802345139

Sanchez-Cortes, D., Aran, O., Jayagopi, D. B., Schmid Mast, M., & Gatica-Perez, D. (2013). Emergent leaders through looking and speaking: from audio-visual data to multimodal recognition. *Journal on Multimodal User Interfaces*, *7*(1), 39–53. https://doi.org/10.1007/s12193-012-0101-0

Sanchez-Cortes, D., Aran, O., Mast, M. S., & Gatica-Perez, D. (2012). A nonverbal behavior approach to identify emergent leaders in small groups. *IEEE Transactions on Multimedia*. https://doi.org/10.1109/tmm.2011.2181941

Schönbrodt, F. D., & Gerstenberg, F. X. R. (2012). An IRT analysis of motive questionnaires: The Unified Motive Scales. *Journal of Research in Personality*, *46*(6), 725–742. https://doi.org/10.1016/j.jrp.2012.08.010

Schönbrodt, F. D., Hagemeyer, B., Brandstätter, V., Czikmantori, T., Gröpel, P., Hennecke, M., … Schulz-Hardt, S. (2018). *Measuring implicit motives with the Picture Story Exercise (PSE): Databases of expert coded german stories, pictures, and updated picture norms*. Manuscript submitted for publication.

Schultheiss, O. C., & Pang, J. S. (2007). Measuring implicit motives. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 322–344). New York, NY: Guilford Press. Retrieved from https://books.google.de/books?hl=en&lr=&id=-VTvN3aPw8sC&oi=fnd&pg=PA322&dq=schultheiss+pang+2007&ots=Fyn8LP3DPD&sig=vi0llCBKQVC89QOJsLn7gjXTi9I

Schultheiss, O. C., & Stanton, S. J. (2009). Assessment of salivary hormones. In E. Harmon-Jones & J. S. Beer (Eds.), *Methods in social neuroscience* (pp. 17–44). New York, NY: Guilford Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=s2XJvHNMolwC&oi=fnd&pg=PA1

7&dq=Assessment+salivary+hormones+Schultheiss+Stanton&ots=kaO0Wo9cKl&sig=xVJyTSj0p5H4BPZAiMyVEj1mDco

Schultheiss, O. C., Schiepe, A., Rawolle, M., Cooper, P. M., Long, D. L., Panter, A. T., & Sher, K. J. (2012). Hormone assays. *APA Handbook of Research Methods in Psychology: Foundations, Planning, Measures, and Psychometrics*, *1*, 489–500. Retrieved from http://www.psych2.phil.fau.de/~oschult/humanlab/publications/SchultheissSchiepeRawolle2012.pdf

Schütz, A., Marcus, B., & Sellin, I. (2004). Die Messung von Narzissmus als Persönlichkeitskonstrukt. *Diagnostica*, *50*(4), 202–218. https://doi.org/10.1026/0012-1924.50.4.202

Schwarzer, R., & Jerusalem, M. (Eds.). (1999). *Skalen zur Erfassung von Lehrer-und Schülermerkmalen: Dokumentation der psychometrischen Verfahren im Rahmen der wissenschaftlichen Begleitung des Modellversuchs selbstwirksame Schulen*. Berlin, Germany. Retrieved from http://www.psyc.de/skalendoku.pdf

Schyns, B., & Paul, T. (2014). Deutsche Self-Monitoring Skala. In D. Danner & A. Glöckner-Rist (Eds.), *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. https://doi.org/10.6102/zis55

Serban, A., Yammarino, F. J., Dionne, S. D., Kahai, S. S., Hao, C., McHugh, K. A., … Peterson, D. R. (2015). Leadership emergence in face-to-face and virtual teams: A multi-level model with agent-based simulations, quasi-experimental and experimental tests. *The Leadership Quarterly*, *26*(3), 402–418. https://doi.org/10.1016/j.leaqua.2015.02.006

Sherer, M., & Adams, C. H. (1983). Construct validation of the Self-Efficacy scale. *Psychological Reports*, *53*(3), 899–902. https://doi.org/10.2466/pr0.1983.53.3.899

Shondrick, S. J., & Lord, R. G. (2010). Implicit leadership and followership theories: Dynamic structures for leadership perceptions, memory, leader-follower processes. *International Review of Industrial and Organizational Psychology*, *25*, 1–33. https://doi.org/10.1002/9780470661628.ch1

Shondrick, S. J., Dinh, J. E., & Lord, R. G. (2010). Developments in implicit leadership theory and cognitive science: Applications to improving measurement and understanding alternatives to hierarchical leadership. *The Leadership Quarterly*, *21*(6), 959–978. https://doi.org/10.1016/j.leaqua.2010.10.004

Smith, J. A., & Foti, R. J. (1998). A pattern approach to the study of leader emergence. *The Leadership Quarterly*, *9*(2), 147–160. https://doi.org/10.1016/S1048-9843(98)90002-9

Smith, J. E., Gavrilets, S., Mulder, M. B., Hooper, P. L., Mouden, C. E., Nettle, D., … Smith, E. A. (2015). Leadership in mammalian societies: Emergence, distribution, power, and payoff. *Trends in Ecology & Evolution*, *31*(1), 54–66. https://doi.org/10.1016/j.tree.2015.09.013

Snyder, M., & Gangestad, S. (1986). On the nature of self-monitoring: Matters of assessment, matters of validity. *Journal of Personality and Social Psychology*, *51*(1), 125–139. https://doi.org/10.1037/0022-3514.51.1.125

Steinmann, B., Dörr, S. L., Schultheiss, O. C., & Maier, G. W. (2015). Implicit motives and leadership performance revisited: What constitutes the leadership motive pattern? *Motivation and Emotion*, *39*(2), 167–174. https://doi.org/10.1007/s11031-014-9458-6

Stogdill, R. M. (1948). Personal factors associated with leadership: A survey of the literature. *The Journal of Psychology*, *25*, 35–71. https://doi.org/10.1080/00223980.1948.9917362

Stogdill, R. M. (1974). Handbook of leadership: A survey of the literature. *The Journal of Psychology*, *25*, 35–71. https://doi.org/10.1080/00223980.1948.9917362

Stueckle, S., & Zinner, D. (2008). To follow or not to follow: decision making and leadership during the morning departure in chacma baboons. *Animal Behaviour*, *75*(6), 1995–2004. https://doi.org/10.1016/j.anbehav.2007.12.012

Stulp, G., Buunk, A. P., Verhulst, S., & Pollet, T. V. (2013). Tall claims? Sense and nonsense about the importance of height of US presidents. *The Leadership Quarterly*, *24*(1), 159–171. https://doi.org/10.1016/j.leaqua.2012.09.002

Stumpf, H., Angleitner, A., Wieck, T., Jackson, D. N., & Beloch-Till, H. (1985). *PRF-D. Personality research form - German version*. Göttingen, Germany: Hogrefe.

Suessenbach, F., Loughnan, S., Schönbrodt, F. D., & Moore, A. B. (2018). The dominance, prestige, and leadership account of social power motives. *European Journal of Personality*, *50*, 7–33. https://doi.org/10.1002/per.2184

Taggar, S., Hackett, R., Saha, S., Hackew, R., Saha, S., Hackett, R., & Saha, S. (1999). Leadership emergence in autonomous work teams: Antecedents and outcomes. *Personnel Psychology*, *52*(4), 899–926. https://doi.org/10.1111/j.1744-6570.1999.tb00184.x

Therneau, T. M. (2015). A package for survival analysis in S (Version 2.43.1) [R package]. Retrieved from https://CRAN.R-project.org/package=survival

Thomas, J. L., Dickson, M. W., & Bliese, P. D. (2001). Values predicting leader performance in the U.S. Army Reserve Officer Training Corps Assessment Center: Evidence for a personality-mediated model. *The Leadership Quarterly*, *12*(2), 181–196. https://doi.org/10.1016/S1048-9843(01)00071-6

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*(5728), 1623–1626. https://doi.org/10.1126/science.1110589

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*, 519–545. https://doi.org/10.1146/annurev-psych-113011-143831

Treffenstaedt, C., & Wiemann, P. (2018). Alfred - A library for rapid experiment development (Version 0.2 b5) [Experimental Software]. Retrieved from https://doi.org/10.5281/zenodo.1437220

Uhl-Bien, M., Riggio, R. E., Lowe, K. B., & Carsten, M. K. (2014). Followership theory: A review and research agenda. *The Leadership Quarterly*, *25*(1), 83–104. https://doi.org/10.1016/j.leaqua.2013.11.007

Umberson, D., & Hughes, M. (1987). The impact of physical attractiveness on achievement and psychological well-being. *Social Psychology Quarterly*, *50*(3), 227–236. https://doi.org/10.2307/2786823

Van der Meij, L., Schaveling, J., & Van Vugt, M. (2016). Basal testosterone, leadership and dominance: A field study and meta-analysis. *Psychoneuroendocrinology*, *72*, 72–79. https://doi.org/10.1016/j.psyneuen.2016.06.005

Van Vugt, M. (2006). Evolutionary origins of leadership and followership. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, *10*(4), 354–371. https://doi.org/10.1207/s15327957pspr1004_5

Van Vugt, M., & Grabo, A. E. (2015). The many faces of leadership: An evolutionary-psychology approach. *Current Directions in Psychological Science*, *24*(6), 484–489. https://doi.org/10.1177/0963721415601971

Van Vugt, M., & Ronay, R. (2014). The evolutionary psychology of leadership theory, review, and roadmap. *Organizational Psychology Review*, *4*(1), 74–95. https://doi.org/10.1177/2041386613493635

Von Collani, G., & Herzberg, P. Y. (2003). Eine revidierte Fassung der deutschsprachigen Skala zum Selbstwertgefühl von Rosenberg. *Zeitschrift Für Differentielle Und Diagnostische Psychologie*, *24*(1), 3–7. https://doi.org/10.1024//0170-1789.24.1.3

Weber, E. U., Blais, A.-R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, *15*(4), 263–290. https://doi.org/10.1002/bdm.414

Weeden, J., & Sabini, J. (2005). Physical attractiveness and health in Western societies: a review. *Psychological Bulletin*, *131*(5), 635–653. https://doi.org/10.1037/0033-2909.131.5.635

Weinert, F. E. (1999). *Konzepte der Kompetenz. Gutachten zum OECD-Projekt: Definition and selection of competencies: Theoretical and conceptual foundations (DeSeCo)*. Neuchatel: Bundesamt für Statistik.

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis (Version 2 3.1.0) [R package]. Retrieved from http://ggplot2.org

Wickham, H., François, R., Henry, L., & Müller, K. (2018). dplyr: A grammar of data manipulation (Version 0.7.7) [R package]. Retrieved from https://CRAN.R-project.org/package=dplyr

Winter, D. G. (1988). The power motive in women—and men. *Journal of Personality and Social Psychology*, *54*(3), 510–519. https://doi.org/10.1037/0022-3514.54.3.510

Winter, D. G. (1994). *Manual for scoring motive imagery in running text*. Ann Arbor: University of Michigan Department of Psychology.

Yukl, G. A. (2013). *Leadership in organizations* (8th ed.). Upper Saddle River, NJ: Prentice Hall. Retrieved from https://books.google.de/books/about/Leadership_in_Organizations.html?id=XFKFygAACAAJ&redir_esc=y

Zaccaro, S. J., Foti, R. J., & Kenny, D. a. (1991). Self-monitoring and trait-based variance in leadership: An investigation of leader flexibility across multiple group situations. *Journal of Applied Psychology*, *76*(2), 308–315. https://doi.org/10.1037/0021-9010.76.2.308

List of Tables

List of Figures

Declaration

Hereby, I confirm that I have written all parts of the thesis by myself, that assistance of third parties was only accepted if scientifically justifiable and acceptable in regards to the examination regulations and that all sources have been quoted.

Göttingen, 30th of April 2019

_____

Johanna Prüfer