The Who, When, and Why of the Glass Cliff Phenomenon

Morgenroth, Thekla; Kirby, Teri A.; Ryan, Michelle K.; Sudkamper, Antonia

Published in:
Psychological Bulletin

DOI:
10.1037/bul0000234

American Psychological Association

Psychological Bulletin

ISSN: 0033-2909

2020, Vol. 146, No. 9, 797–829
http://dx.doi.org/10.1037/bul0000234

The Who, When, and Why of the Glass Cliff Phenomenon: A Meta-Analysis of Appointments to Precarious Leadership Positions

Thekla Morgenroth and Teri A. Kirby
University of Exeter

Michelle K. Ryan
University of Exeter and University of Groningen

Antonia Sudkämper
University of Exeter

Women and members of other underrepresented groups who break through the glass ceiling often find themselves in precarious leadership positions, a phenomenon that has been termed the *glass cliff*. The glass cliff has been investigated in a range of domains using various methodologies, but evidence is mixed. In 3 meta-analyses, we examined (a) archival field studies testing whether members of underrepresented groups, compared with members of majority groups, are more likely to be appointed to leadership positions in times of crisis; (b) experimental studies testing whether members of underrepresented groups, compared with members of majority groups, are evaluated as more suitable for, as well as (c) more likely to be selected for, leadership positions in times of crisis. All 3 analyses provided some evidence in line with the glass cliff for women. Specifically, the meta-analysis of archival studies revealed a small glass cliff effect that was dependent on organizational domain. The leadership suitability meta-analysis also showed a small glass cliff effect in between-participants studies, but not in within-participants studies. The analysis of leadership selection revealed that women are more likely to be selected over men in times of crisis, and that this effect is larger in countries with higher gender inequality. The glass cliff also extended to members of underrepresented racial and ethnic groups. We explore several moderating factors and report analyses shedding light on the underlying causes of the glass cliff. We discuss implications of our findings as well as open questions.

Public Significance Statement
These meta-analyses demonstrate a small effect in line with the glass cliff; that is, women and members of underrepresented racial and ethnic groups are more likely to be rated as suitable for, and appointed to, leadership positions in times of crisis. Yet, these effects are context-dependent, reinforcing the nuanced nature of the glass cliff phenomenon.

Keywords: glass cliff, leadership, management, meta-analysis, women in leadership

Whereas men remain overrepresented in positions of power and influence, in recent years there has been slow but steady change, as we see women breaking through the glass ceiling. For example, whereas women only made up 9.5% of Fortune 500 company boards in 1995 (Catalyst, 1995) this number has risen to 22.5% in 2018 (Catalyst, 2019). Similarly, the number of women on company boards in Europe increased from 13.9% in 2011 to 33.6% in 2018 (EWOB, 2018). Similar trends can be seen in politics, where

This article was published Online First July 23, 2020.

Thekla Morgenroth and Teri A. Kirby, Department of Psychology, University of Exeter; Michelle K. Ryan, Department of Psychology, University of Exeter, and Faculty of Economics and Business, University of Groningen; Antonia Sudkämper, Department of Psychology, University of Exeter.

Antonia Sudkämper is now at Cambridge Asssessment, Cambridge, United Kingdom.

Thekla Morgenroth and Teri A. Kirby contributed equally to the article and are joint first authors. Thekla Morgenroth uses they/them/their pronouns, and Teri A. Kirby, Michelle K. Ryan, and Antonia Sudkämper use she/her/hers pronouns.

The data sets, analysis code, and codebook for analyses are available at https://osf.io/b8tzq/?view_only=6aad2cc8f1ba4041bcd2cc48e44cb4aa.

The authors thank Alex Haslam for his helpful feedback on this article, Leire Gartzia, Helix Grosse-Stolentenberg, Max Montgomery, and Paweena Orapin for their help with the coding of reports in languages other than English, and Clara Tucker for her help with the reference list. We also thank all authors who shared their unpublished data with us or provided additional information about their studies. This work was supported by a European Commission Grant (725128) awarded to Michelle K. Ryan and a New Investigator Grant awarded to Teri A. Kirby by the Economic and Social Research Council (Grant ES/S00274X/1).

Correspondence concerning this article should be addressed to Thekla Morgenroth, Department of Psychology, University of Exeter, WA Singer Laboratories, Perry Road, Exeter EX4 4QG, United Kingdom. E-mail: t.morgenroth@exeter.ac.uk

the number of women in national parliaments worldwide has risen from 11.69% in 1997 to 24.58% in 2019 (The World Bank, 2019).

Despite these gains, research suggests that those women who do break through the glass ceiling may be more likely to find themselves on a *glass cliff*—such that their leadership position can be seen as being relatively risky or precarious compared with that of their male counterparts (Ryan & Haslam, 2005, 2007). For example, research suggests that women are disproportionately likely to be appointed to leadership positions in times of crisis, whether that be an organizational crisis or a political one. The term *glass cliff* was coined by Ryan and Haslam (2005), who described the phenomenon in response to an article in *The Times* (Judge, 2003), claiming that women who broke through the glass ceiling and took on leadership positions in U.K. boardrooms had "wreaked havoc on companies' performance" (p. 21). Ryan and Haslam reanalyzed the same data that the newspaper article reported and came to a very different conclusion. Rather than *causing* a drop in stock performance, as Judge had claimed, it appeared that women were more likely than men to be appointed as board members after companies had already experienced a sustained pattern of poor share price performance.

Since this first identification of the phenomenon, researchers have investigated the glass cliff in many different ways. It has been examined across a range of domains such as management (e.g., Haslam & Ryan, 2008), politics (e.g., Kulich, Ryan, & Haslam, 2014), legal work (Ashby, Ryan, & Haslam, 2007), sports (Wicker, Cunningham, & Fields, 2019), and education (Smith, 2015). The glass cliff has also been investigated across various countries, including the U.K. (e.g., Haslam & Ryan, 2008), the US (e.g., Bruckmüller & Branscombe, 2010), Canada (Hennessey, MacDonald, & Carroll, 2014), Germany (e.g., Bechtoldt, Bannier, & Rock, 2019), The Netherlands (e.g., Rink, Ryan, & Stoker, 2012); Switzerland (Kulich, Iacoviello, & Lorenzi-Cioldi, 2018); and Turkey (Acar, 2015; Kurt, 2011; Uyar, 2011).

Research into the glass cliff has also utilized a range of methodologies, including archival studies (e.g., Bechtoldt et al., 2019; Haslam, Ryan, Kulich, Trojanowski, & Atkins, 2010), qualitative research (e.g., Peterson, 2016; Wilson-Kovacs, Ryan, Haslam, & Rabinovich, 2008), panel surveys (e.g., Sabharwal, 2015), and experiments (e.g., Ashby et al., 2007; Bruckmüller & Branscombe, 2010; Haslam & Ryan, 2008; Rink, Ryan, & Stoker, 2013). Additionally, some studies have explored the possibility that the glass cliff may also affect other groups, such as underrepresented ethnic and racial groups (e.g., Cook & Glass, 2013, 2014a; Gündemir, Carton, & Homan, 2019; Kulich et al., 2014).

Across all of these studies, evidence for a glass cliff is mixed. Whereas some find evidence for the phenomenon (e.g., Haslam & Ryan, 2008; Ashby et al., 2007; Kulich et al., 2014), others do not (e.g., Adams, Gupta, & Leeth, 2009; Bechtoldt et al., 2019; Hennessey et al., 2014). In line with Ryan and Haslam's (2009) reasoning that the glass cliff is not a universal phenomenon (as we will discuss in further below; see also Ryan et al., 2016), still others find evidence for the glass cliff only under specific circumstances (e.g., Bruckmüller & Branscombe, 2010; Kulich, Lorenzi-Cioldi, Iacoviello, Faniko, & Ryan, 2015).

Although a number of review articles have examined variability in the glass cliff phenomenon (e.g., Ryan & Haslam, 2009; Ryan et al., 2016), this sense-making has thus far been restricted to theoretical analysis. To our knowledge, no prior article has provided a systematic review and meta-analysis of the glass cliff literature. In the present review, we fill this gap by conducting three separate meta-analyses exploring key aspects of the glass cliff phenomenon. More specifically, we meta-analyze (a) archival field studies that test whether women are more likely than men to be appointed to leadership positions in times of crisis, (b) experimental studies investigating whether women are evaluated more positively than men in times of crisis (vs. when all is going well), and (c) experimental studies investigating whether women are more likely than men to be selected for leadership positions in times of crisis (vs. when all is going well). We also investigate whether members of underrepresented racial and ethnic groups also encounter the glass cliff.

The Glass Cliff

Since the initial demonstration of the glass cliff in a relatively small sample (Ryan & Haslam, 2005), researchers have replicated the phenomenon several times in the context of corporate management. For example, Cook and Glass (2014a) investigated the glass cliff in the US and found that poorly performing Fortune 500 companies were more likely to appoint a female CEO than those performing well. Similarly, Haslam and colleagues (2010) investigated the glass cliff in FTSE 100 companies in the years 2001–2005 and found evidence that stock performance was negatively related to the presence of women on company boards the following year. In other words, the worse the performance, the higher the likelihood that the company board would have at least one female member the next year. However, other researchers investigating the glass cliff in the management domain have not found evidence for the phenomenon and have called its existence into question (e.g., Adams et al., 2009; Bechtoldt et al., 2019).

In addition to the management domain, the glass cliff has been examined in a range of other contexts. For example, Ryan, Haslam, and Kulich (2010) found that in the political domain, at least within the U.K. Conservative Party, women were more likely to contest parliamentary seats that were currently held by the opposite party by a greater margin; that is, those for which the risk of failure (i.e., not getting elected) was higher (see also Kulich et al., 2014). Moreover, using a sample of educational agencies, Smith (2015) found that educational leaders were more likely to be female than male when the risk of failure was higher, as measured by indicators such as the percentage of students with limited English knowledge and out-of-school suspensions. These findings indicate that the glass cliff is not restricted to a specific context or specific measures but may occur under many circumstances.

Evidence for the glass cliff extends to experiments as well. Such findings suggest a causal path from company performance to the appointment of female leaders. For example, across three studies, Haslam and Ryan (2008) presented management graduates, high-school students, and business leaders with fictitious information about a company that was either performing well or poorly and asked them to select a leader. They found evidence for the glass cliff, such that participants (a) saw a female candidate as having higher leadership ability compared with a male candidate and (b) were more likely to select a female leader in the context of poor performance compared with when performance was strong (see also Ashby et al., 2007; Brown, Diekman, & Schneider, 2011; Hunt-Earle, 2012).

What Are the Underlying Mechanisms of the Glass Cliff Phenomenon?

Researchers have suggested and examined a range of potential explanations for the glass cliff, with suggestions that the glass cliff is complex and multiply determined (e.g., Ryan & Haslam, 2007; Ryan et al., 2016). One reason why individuals and organizations may prefer women in times of crisis is societally shared gender stereotypes, and the resulting match (or mismatch) between how men and women are perceived and perceptions of the attributes that are needed for effective leadership in different situations. Research shows that women are seen as being more likely to possess communal traits, such as being cooperative and caring, whereas men are seen as more likely to possess agentic traits such as being assertive and independent (e.g., Eagly & Wood, 2012). In turn, men are generally seen as a better fit for leadership positions, which are thought to require agentic traits, the *think manager—think male* association (Heilman, 2001; Schein, 1973).

However, Ryan, Haslam, Hersby, and Bongiorno (2011) found that this association between masculinity and leadership reverses in times of crisis. In other words, when times are difficult, stereotypically feminine characteristics are seen as important traits for leaders to have, what the authors call a *think crisis—think female* association. In turn, women are perceived to have a better fit with leadership positions and are more likely to be chosen as leaders (Bruckmüller & Branscombe, 2010). Thus, if women are indeed appointed to leadership positions in times of crisis because of the *think crisis—think female* association, it should be restricted to cases in which stereotypically feminine attributes are needed (e.g., social or emotional skills), not when stereotypically masculine attributes are needed (e.g., technical expertise or ruthlessness; e.g., Ryan et al., 2011; Schürmann, 2017). We will test this prediction.

Because these stereotypes are specifically associated with women, it seems unlikely that this process would generalize equally to other underrepresented groups who have been shown to face a glass cliff. But, to the extent that group stereotypes overlap with feminine gender stereotypes, similar processes may be involved here too. For example, Galinsky, Hall, and Cuddy (2013) found that Asian stereotypes are more feminine than White stereotypes, while Black stereotypes are more masculine. Indeed, there is initial evidence that the *think crisis—think female* association may be extended to some racial and ethnic groups, such as East Asian Americans (e.g., Gündemir et al., 2019).

Organizations may also appoint members of underrepresented groups in times of crisis to signal change by shifting away from previous leadership choices (i.e., white men). In other words, when things are going badly—particularly in terms of negative reactions from shareholders or the public more widely—organizations may wish to communicate that they are taking a new approach. Appointing a leader who is different from previous leaders might be an effective way of signaling this change, and evidence supports this explanation. For example, Kulich and colleagues (2015) demonstrated that participants' *beliefs* that a female candidate's ability to signal change, but not her *ability* to be effective in the position, explained the preference for the female candidate in times of crisis. A moderator speaking to this explanation is the history of leadership in an organization. Some experimental studies have manipulated the gender of previous leaders, arguing that women make a poor signal for change if the previous leaders were already female. Findings support this argument (Bruckmüller & Branscombe, 2010), and we will include history of leadership as a moderator in our analyses. Given that this explanation is not tied to gender-specific stereotypes, we suggest that this process should also occur for members of other groups that are underrepresented in leadership positions such as members of some racial and ethnic groups.

Finally, it has been suggested that the glass cliff is the result of prejudice. In other words, women are appointed to precarious leadership positions because such positions are less desirable and may set them up for failure (see Ryan et al., 2011, Study 3; Ryan et al., 2016). Accordingly, evidence suggests that the glass cliff is more likely to occur among conservatives (Ryan et al., 2010) and individuals high in legitimizing ideology (Brown et al., 2011). Additionally, findings suggest that the glass cliff is more likely to occur when organizational stakeholders do not support the appointment of a new leader, supporting the idea that women are set up to fail (Rink et al., 2013). Similar to the *signaling change* explanation, prejudice could also explain the emergence of the glass cliff among other underrepresented groups such as racial and ethnic minorities.

Importantly, these three explanations are not mutually exclusive (Ryan et al., 2016). It could well be that all three of these processes feed into the emergence of the glass cliff. We will investigate the extent to which each of the three explanations is supported in our analyses.

Overview and Predictions

The glass cliff has been investigated using both archival and experimental methods. Archival methods are, by nature, correlational and vary more widely in the type of measures they use, but they have higher external validity. Experimental methods are generally more similar to each other in their approach and measures and allow causal inferences. In the present analysis, we will investigate the presence of the phenomenon separately for these two approaches. For experimental studies, we will conduct one meta-analysis on the *evaluation* of female and male candidates in terms of suitability or leadership ability (in both within-participants and between-participants designs) and one meta-analysis on the *selection* of female and male candidates (only available in within-participants designs). We will examine these separately because evaluations of women's leadership and their subsequent selection (or not) for a leadership position do not always map onto each other. For example, people sometimes hold lower minimum standards for women than for men when rating their abilities or considering them for an interview but can hold higher standards for women when selecting them for positions (Biernat & Fuegen, 2001).

For archival studies, we predict that there will be more organizational struggles (such as, but not limited to, negative performance) preceding the appointment of women to leadership positions compared with the appointment of men. For the experimental studies, we predict that female candidates will be more likely to be selected and evaluated more positively relative to male candidates in times of crisis compared with when no crisis exists. This can include (a) female leaders being evaluated more positively than men overall, but particularly in times of crisis, (b) male leaders being evaluated more positively than women overall, but less so in times of crisis, or (c) male leaders being evaluated more positively

in times of success, but female leaders being evaluated more positively in times of crisis. In other words, we have no prediction as to whether there will be an overall preference for female or male leaders, but we do predict that the preference for one gender over the other (or lack thereof) will shift more toward preferring women in times of crisis compared with when no crisis exists. To conduct these analyses in the experimental studies, we will compare the effect size for women relative to men in crisis and no-crisis conditions. For all three meta-analyses, we will test whether these same effects also extend to members of underrepresented racial and ethnic groups.

Moderators

We will examine several moderating factors to explore the *who*, *when*, and *why* of the glass cliff. Although none of the moderators can shed light on the *why* directly, the emerging pattern of results can indirectly lend support to the *think crisis—think female*, *signaling change*, or *prejudice* explanations.

Moderators of Theoretical Interest

For the *archival studies*, we will include three moderators that can speak to the explanations put forward for the glass cliff. First, we will test whether the strength of the glass cliff effect varies by domain. If the effect is larger in particularly male-dominated domains, then this finding may be interpreted as support for the notion that women are appointed to glass cliff positions to signal change, given that their appointment is a stronger signal if it is a rarer occurrence. It could also lend support for the *prejudice* explanation, given that gender bias is generally larger in male-dominated fields.

Second, we will test whether the strength of the glass cliff differs by target (women vs. underrepresented racial and ethnic groups). If the effect is larger (or only present) for women, this finding would support the *think crisis—think female* explanation (i.e., that women are appointed because stereotypical female attributes are seen as useful in times of crisis). Lastly, we will test whether gender inequality in different countries in different years moderates the effect. If we find that the glass cliff is more pronounced in countries with higher levels of gender inequality, then it would support the argument that prejudice underlies the glass cliff phenomenon.

For experimental studies, we will again include several moderators examining three of the proposed causes of the glass cliff. First, we will test whether history of leadership (previous leader was male vs. female) moderates the effect. If the glass cliff effect is larger (or only present) when the previous leader was male, then we will interpret this finding as support for the notion that women are appointed in times of crisis to signal change. Next, we will investigate whether type of crisis (feminine vs. masculine) moderates the effect. If the glass cliff effect is larger (or only present) when the crisis is feminine (i.e., when it requires stereotypically feminine attributes such as communication skills), then it would support the *think crisis—think female* explanation. Similar to the archival analysis, we will test whether the glass cliff equally applies to underrepresented ethnic and racial groups. If the effect is larger for studies examining target gender, then we will interpret this finding as support for the *think crisis—think female* explanation. If the effect does not differ between targets, then it would support the *signaling change* and the *prejudice* explanations.

To further explore evidence for or against the three main explanations for the glass cliff, we will also examine the magnitude of the effects across different racial and ethnic groups. If racial or ethnic groups more associated with feminine stereotypes (e.g., East Asian individuals) demonstrated a larger glass cliff than those associated with masculine stereotypes (e.g., Black individuals; see Galinsky et al., 2013), then it would provide evidence that the glass cliff occurs because stereotypically feminine characteristics are seen as useful in times of crisis. If not, then our results would be more in line with the possibility that organizations in crisis are motivated to signal that they are embracing change or that the decision is based on gender and racial biases.

We will further examine support for the *prejudice* explanation by testing whether gender inequality in different countries in different years (measured by the gender inequality index) and gender of participants (measured as percentage of male participants in the sample) moderate the glass cliff effect. If the size of the effect is larger in samples with a higher proportion of male participants, then it would suggest that ingroup favoritism (i.e., a form of bias) might play a role in the appointment of women to glass cliff positions. If the glass cliff is more pronounced in countries with higher gender inequality, then it would also support the notion that the glass cliff is a reflection of a wider pattern of gender prejudice.

Study Quality, Methodology, and Bias

We include several moderators in our analyses to investigate the role that study quality, methodological differences, and bias may play in the extent to which authors have found the glass cliff. To test for potential bias, we will test whether the involvement of at least one of the original authors (i.e., Ryan or Haslam) in the study moderates the effect. One of the original authors (the third author of the current review) has been involved in a considerable number of glass cliff studies, and we consider it worthwhile to test whether the effect replicates in studies in which the original authors were not involved. We will also test whether the effect is larger for published studies compared with unpublished studies as part of our publication bias analysis.

To examine whether study quality affects the emergence of the glass cliff, we include several quality moderators. For the archival studies, indicators of high quality are: (a) type of appointment measure, that is, whether the study reports the *change* in the number of underrepresented candidates rather than their static numbers (e.g., appointment of women to boards vs. presence of women on boards); (b) type of performance measure, that is, whether the performance measure indicates *change in performance over time* rather than *static* performance (e.g., change in stock price vs. stock price); (c) the time between the two measures, that is, whether the measure of the performance indicator was no more than a year before the appointment (otherwise, it is less plausible that the appointment was indeed made in response to the performance); and (d) researcher experience, that is, whether the study was carried out by an academic (including doctoral-level students) or a student below Ph.D. level.

For the experimental studies, indicators of high quality include (a) whether the study successfully manipulated the performance of

the organization, (b) whether the dependent variable showed high reliability, and (c) researcher experience, that is, whether the study was carried out by an academic (including doctoral-level students) or a student below Ph.D. level.

Finally, we will test whether methodological differences affect the size of the glass cliff effect. For archival studies in the management domain, we will test whether the effect size differs for accountancy-based versus stock-based measures. For experimental studies, we will test whether type of participants (undergraduate students vs. nonundergraduate sample) moderates the effect. A larger glass cliff effect in nonundergraduate samples could suggest that workplace experience (e.g., working with female and male managers) plays a role in the emergence of the phenomenon.

Method

Literature Search

Two of the authors independently searched the literature to locate relevant studies. The most recent search of the literature was conducted in March, 2019. We searched Google Scholar, PsycARTICLES, PsycINFO, ERIC, ProQuest Business Premium Collection, and Business Source Complete to locate relevant reports. To locate unpublished research such as dissertations and conference papers, we also searched the ProQuest Dissertation & Theses Global database, PsycEXTRA, and ETHOS. We searched for the term "glass cliff," and, to ensure that we were not missing any studies that did not use this term but nevertheless investigated the phenomenon, we additionally searched for (crisis OR "organizational performance" OR "company performance" OR threat) AND ("leader selection" OR "leader appointment" OR "leader preference"). To determine whether an article was relevant, we independently read the title, abstract, or both, and compared our results. Any discrepancies were resolved through discussion.

We then took additional steps to find published studies we might have missed, but particularly unpublished studies to address potential publication bias. First, we distributed announcements requesting unpublished studies on the glass cliff effect through the listservs or websites of six relevant organizations: (a) British Academy of Management, (b) European Academy of Management, (c) European Association of Social Psychology, (d) European Association of Work and Organizational Psychology, (e) Society for Australasian Social Psychology, and (f) Society for Personality and Social Psychology. We made these announcements twice—once shortly after our first literature search (in 2015) and once at the time of our third literature search (in 2018).

Moreover, we e-mailed at least one of the authors of each report identified in the first literature search to ask whether they had conducted any additional research on the glass cliff. In most cases, we contacted the first author. In the case of undergraduate and master's dissertations and theses, we contacted the primary advisor. Finally, we contacted the authors of several studies presented at conferences.

Inclusion and Exclusion Procedures

In line with preregistered specifications (https://osf.io/j7t46?view_only=6aad2cc8f1ba4041bcd2cc48e44cb4aa), to be included in our meta-analyses, reports had to meet three criteria. First, they had to report at least one quantitative study in which the relationship between organizational performance (including the presence of some form of crisis or threat) and preference for members of underrepresented groups for a leadership role was examined. *Preference* in this context refers to the selection as well as evaluation of members of underrepresented groups. Second, the data had to report sufficient information to calculate the necessary statistical information. This information could be provided in the text itself or by one of the authors or calculated from data sets shared by the authors. Finally, the data had to be independent, that is, not contained in another report included in the meta-analysis.

For experimental studies, the independence criteria were relatively straightforward. For example, there were cases in which data were first reported in a student dissertation and later published in an academic journal. In these cases, the most recent time the data were reported was included.[1] For archival studies, the issue of independence was more complex. There were several instances in which different groups of authors had analyzed data from the same organizations in overlapping years, using slightly different measures of company performance and preference for majority or minority candidates. For example, Elsaid and Ursel (2018) investigated the glass cliff in CEO appointments of S&P 1000 companies between 1992 and 2014, and Oelbaum (2016) investigated the glass cliff in CEO appointments of Fortune 500 companies between 2000 and 2014. Whereas these samples and years are not exactly the same, they overlap significantly. Including both reports in the meta-analysis would thus mean that any companies listed in both the Fortune 500 and the S&P 1000 between 2000 and 2014 would be included twice in the meta-analysis. In cases like these, we based the decision of which study to include on four criteria. First, in line with the quality criteria discussed above, if one study used change in number of women or members of underrepresented racial or ethnic groups (e.g., number of women *appointed* to the board of directors) while the other(s) did not (e.g., *presence* of women on board of directors), we included the one examining the change in numbers as it more closely matches the definition of the glass cliff. Second, if one study used change of company performance over time (e.g., *change* in stock price) while the other(s) used static measures (e.g., stock price), we included the one reporting change in performance. Third, we preferentially included studies that contain a higher number of relevant data points (e.g., number of female leaders appointed), for example, by spanning a longer period of time or including a larger sample of organizations. Fourth, we preferentially included studies for which the time between company performance and appointment/presence of underrepresented groups was a year or less over studies where the studied time was more than a year. The same criteria were used when deciding between measures within the same study, with the exception of the third criterion—multiple measures could instead be combined, and the sample sizes were nearly always identical.

[1] In one case (Bruckmüller & Branscombe, 2010), some of the necessary statistical information was not available, so we coded the information from the earlier, unpublished version (Bruckmüller, 2007).

When it was possible to include only the nonoverlapping data from two studies (e.g., because the authors reported associations separately for each year), we excluded the overlapping data from one study and include both studies. When the overlap of data between two studies was less than 10% for both studies, we included both studies to maximize the data we were able to include. Table 4 lists all studies that were excluded as well as the reason for exclusion.

The final overall sample included 91 independent samples from 74 independent studies (see Figure 1 and Tables 1, 2, and 3), 58 of which came from experimental studies and 33 of which came from archival studies. Some studies included moderators that were relevant to our meta-analysis, so we included those effects separated by the relevant condition, which meant that a single study could contribute more than one effect size. For example, if a between-participants experiment manipulated whether the nature of the crisis required stereotypically feminine or masculine attributes, we calculated a separate effect size for the feminine and masculine conditions. Because all of these moderators were manipulated between participants, the effect sizes were still independent in the sense that the participants providing the data were not overlapping. At the same time, some effect sizes were dependent because they were part of the same study and therefore were conducted in similar contexts.

Coding of Variables

Two independent coders coded the variables reported here using detailed coding guidelines (see https://osf.io/qn423/?view_only=48dc9f5a0ecc4c43b942440392886487) for all studies available in English or German, based on the language skills of the authors. For reports in other languages (i.e., Norwegian, Spanish, and Thai), the most experienced coder coded the report together with a native speaker of the language in question. These native speakers included a psychology Master's student, a management doctoral-level student, and a social psychology academic. The coding process for these studies involved the coder reading the report using online translation software and the native speaker reading the report before a meeting. During the meeting, the coder explained in detail the information we needed, confirmed information obtained with the help of the online translation, and answered any questions that the native speaker may have had.

Coding of archival studies. In the archival studies, authors generally reported the relationship between an indicator of organizational performance as the predictor variable and the group membership (gender, race/ethnicity) of a leader as the outcome variable. This relationship was either reported as a correlation (e.g., the correlation between the appointment of a man vs. woman and the average share price in the previous year) or as mean

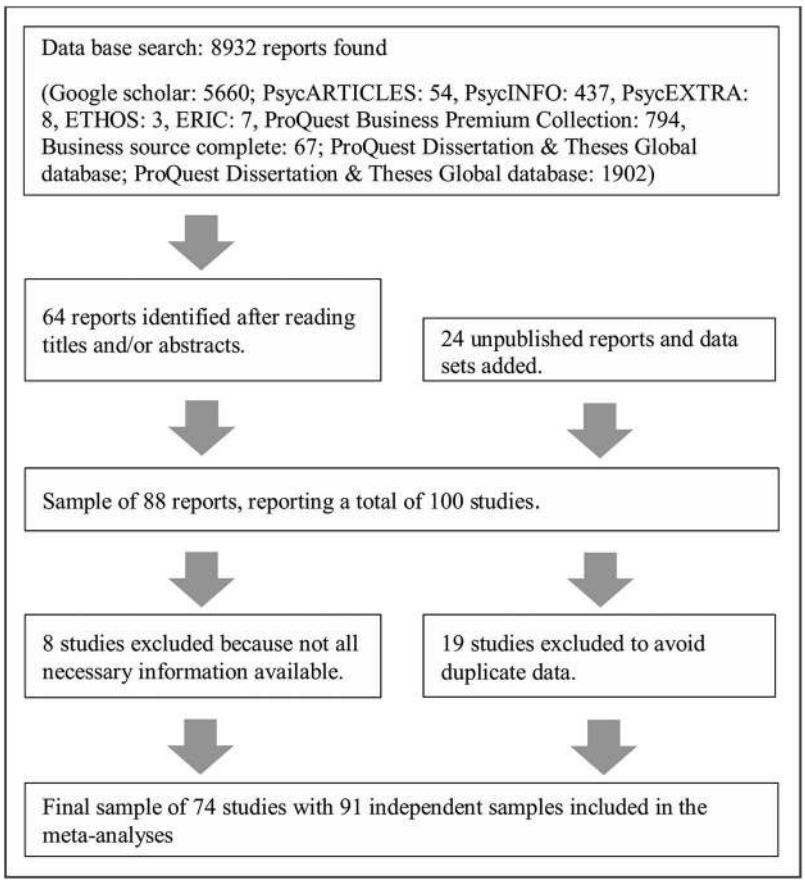

Figure 1. Flowchart of studies included in the meta-analyses.

Table 1
Overview of Reports Included in the Meta-Analysis of Archival Studies

Report	Domain	Target characteristic	GII	Original authors	Published	Appointment measure	Performance measure	Lag ≤ 1 year	Researcher experience	Performance measure(s)	*g*	*N*
Ahlden & Kollberg, 2018	Management	Gender	.046	No	No	Appointment of female leaders	Change in performance	Yes	Student	Tobin's Q, EBIT, RoA, RoE	0.04	408–456
Bechtoldt et al., 2019	Management	Gender	.170	No	Yes	Appointment of female leaders	Static performance	Yes	PhD student or above	RaR, RoA, RoE	−0.04	226
Bowles, 2013	Education	Gender	.222	No	No	Presence of female leaders	Static performance	Yes	PhD student or above	Performance of school, poverty of students	0.43	83
Browning, 2017	Politics	Gender	.253	No	No	Appointment of female leaders	Static performance	No	Student	% margin of previous electoral success/failure	−0.06	1602
Cook & Glass, 2013	Sports	Race/Ethnicity	—	No	Yes	Appointment of racial/Ethnic minority leaders	Static performance	Yes	PhD student or above	Winning record in previous season	0.10	1130
Cook & Glass, 2014b	Management	Race/Ethnicity	—	No	Yes	Appointment of racial/Ethnic minority leaders	Static performance	Yes	PhD student or above	Shareholder return, RoA, RoE	−0.07	551–591
Dula, 2018	Non-profit	Gender	.235	No	No	Presence of female leaders	Static performance	Yes	PhD student or above	Allocated budget	0.49	1580
Funk, 2017	Politics	Gender	.457	No	No	Appointment of female leaders	Static performance	No	PhD student or above	Past vote share	0.04	41114
Haslam, Ryan, Kulich, Trojanowski, & Atkins, 2010	Management	Gender	.215	Yes	Yes	Presence of female leaders	Static performance	Yes	PhD student	Tobin's Q, RoA, RoE (2002)	−0.40	92–95
			.206						or above	Tobin's Q, RoA, RoE (2004)	−0.03	90–92
Hennessey et al., 2014	Management	Gender	.132	No	Yes	Appointment of female leaders	Change in performance	Yes	PhD student or above	CER (2005)	−0.08	38
			.121							CER (2007)	−0.07	24
Ihmels, Shemla, & Wegge, 2019a	Management	Gender	.103	No	No	Appointment of female leaders	Change in performance	Yes	PhD student	Share price trend (Study 1)	0.08	126
			.099						or above	Share price trend (Study 2)	0.07	84
Kleineberg, 2018	Politics	Gender	.078	No	No	Appointment of female leaders	Static performance	No	Student	% margin of previous electoral success/failure	0.45	299

(*table continues*)

Table 1 (*continued*)

Report	Domain	Target characteristic	GII	Original authors	Published	Appointment measure	Performance measure	Lag ≤ 1 year	Researcher experience	Performance measure(s)	*g*	*N*
Kulich, Ryan, & Haslam, 2014	Politics	Gender	.215	Yes	Yes	Appointment of female leaders	Static performance	No	PhD student or above	% margin of previous electoral win/loss (2001)	0.04	1280
			.206							% margin of previous electoral win/loss (2005)	0.09	1167
			.172							% margin of previous electoral win/loss (2010)	0.06	1260
		Race/Ethnicity	—	Yes	Yes	Appointment of racial/Ethnic minority leaders	Static performance	No	PhD student or above	% margin of previous electoral win/loss (2001)	0.26	1280
										% margin of previous electoral win/loss (2005)	0.60	1167
										% margin of previous electoral win/loss (2010)	0.25	1259
Mile & Undheim, 2018	Management	Gender	.082	No	No	Appointment of female leaders	Static performance	Yes	Student	RoA, RoE (previous leader male)	−0.12	306
										RoA, RoE (previous leader female)	−0.65[a]	16
Oelbaum, 2016	Management	Gender	.260	No	No	Appointment of female leaders	Change in performance	Yes	PhD student or above	Shareholder return, RoA, RoE, ROIC-WACC	−0.14	70–80
Robinson, 2019	Politics	Gender	.207	No	No	Appointment of female leaders	Static performance	No	Student	% margin of previous electoral win/loss (House)	0.15	8072
										% margin of previous electoral win/loss (Senate)	0.07	2282
Ryan & Haslam, 2005	Management	Gender	.211	Yes	Yes	Appointment of female leaders	Change in performance	Yes	PhD student or above	Share price change	−0.01	38
Smith, 2014	Management	Gender	.305	No	Yes	Presence of female leaders	Static performance	Yes	PhD student or above	Risk (Consequences of Failure × Likelihood of Failure)	−0.17	142

Table 1 (*continued*)

Report	Domain	Target characteristic	GII	Original authors	Published	Appointment measure	Performance measure	Lag ≤ 1 year	Researcher experience	Performance measure(s)	*g*	*N*
Smith, 2015	Education	Gender	.264	No	Yes	Presence of female leaders	Static performance	No	PhD student or above	Complexity of work (% gifted students, % students with limited English proficiency, % of out of school suspensions)	0.23	911
Smith & Monaghan, 2013	Management	Gender	.237	No	Yes	Presence of female leaders	Static performance	Yes	PhD student or above	Risk (Consequences of Failure × Likelihood of Failure)	0.15	118
Sun, Zhu, & Ye, 2015	Management	Gender	.210	No	Yes	Presence of female leaders	Change in performance	Yes	PhD student or above	Investment growth	−0.01	14609
Weber, 2016	Management	Gender	.090	No	No	Appointment of female leaders	Change in performance	Yes	Student	Tobin's Q change	−0.03	906
Wicker, Cunningham, & Fields, 2019	Sports	Gender	.227	No	Yes	Appointment of female leaders	Static performance	Yes	PhD student or above	number of losses in previous season, number of wins in previous season. % of games won in previous season	0.15	59

Note. *g* = Hedges' *g* (bias-corrected) where positive values indicate higher appointment of women or minorities in crisis situations; GII = gender inequality index (only relevant for studies investigating the glass cliff for women); CER = cumulative excess returns; EBIT = earnings before interest and taxes; RoA = return on assets; RoE = return on equity; ROIC-WACC = return on invested capital minus weighted average cost of capital.

[a] RoA was not included in the calculation for *g* because it was an extreme outlier (more than four standard deviations below the mean). The value with RoA included is $g = -1.40$.

Table 2
Overview of Reports Included in the Leadership Suitability Meta-Analysis

Report	Study	Target characteristic	Type of crisis	Gender of previous leader(s)	Participant gender (%female)	Undergraduate sample	GII	Published	Original authors	Researcher experience	Manipulation check	DV reliable	*g*	*N*
Acar, 2015		Gender	Neutral/unclear	Unknown	49.82	Yes	.348	No	Yes	PhD student or above	No	Yes	0.01	281
Acar & Sümer, 2018		Gender	Neutral/unclear	Unknown	52.69	No	.316	No	Yes	PhD student or above	Yes	Yes	0.05	373
Ashby, Ryan, & Haslam, 2007		Gender	Feminine	Unknown	66.67	Yes	.206	Yes	Yes	PhD student or above	Yes	Yes	0.38	113
Bain et al., 2011		Gender	Neutral/unclear	Unknown	65.38	Yes	.129	No	No	PhD student or above	Yes	Yes	0.19	130
Bain & Bongiorno, 2011		Gender	Neutral/unclear	Unknown	31.48	Yes	.129	No	No	PhD student or above	Yes	Yes	−0.18	95
Boonsom, 2013		Gender	Neutral/unclear	Female Male	48.00	No	.317	No	No	Student	No	Yes	0.08 0.27	96 95
Brown, Diekman, & Schneider, 2011	3	Gender	Masculine	Unknown	43.18	Yes	.242	No	Yes	PhD student or above	Yes	Yes	1.05	87
	4	Gender	Masculine	Unknown	66.12	Yes	.242	No	Yes	PhD student or above	Yes	Yes	0.85	120
Bruckmüller & Branscombe, 2010	2	Gender	Neutral/unclear	Unknown	50.00	Yes	.242	No	Yes	PhD student or above	Yes	Yes	0.64	109
Bruckmüller, Barthel, Bzduch, Dreisbach, & Straubmeier, 2008		Gender	Neutral/unclear	Unknown	62.96	Yes	.092	No	No	Student	Yes	Yes	0.28	53
Chambers, 2011		Gender	Neutral/unclear	Unknown	59.45	No	.113	No	No	Student	No	No	0.00	289/290
Gündemir, Carton, & Homan, 2019	2	Race/Ethnicity	NA	NA	51.98	No	NA	No	No	PhD student or above	Yes	No	0.40	199
	4	Race/Ethnicity	NA	NA	45.23	No	NA	No	No	PhD student or above	Yes	Yes	0.10	227
Haslam & Ryan, 2008	1	Gender	Neutral/unclear	Unknown	65.59	Yes	.206	Yes	Yes	PhD student or above	No	Yes	0.34	91/93
	2	Gender	Feminine	Unknown	71.76	Yes	.206	Yes	Yes	PhD student or above	No	No	0.16	85
	3	Gender	Neutral/unclear	Unknown	56.58	No	.206	Yes	Yes	PhD student or above	No	Yes	1.13	83
Hirschfeld, Ihmels, & Wegge, 2018	1	Gender	Neutral/unclear	Unknown	58.00	No	.072	No	No	Student	Yes	Yes	0.10	200
Ihmels, Shemla, & Wegge, 2017	1	Gender	Neutral/unclear	Female Male	45.31	No	.072	No	No	PhD student or above	Yes	Yes	−0.10 0.19	95 94
	2	Gender	Neutral/unclear	Unknown	59.55	Yes	.072	No	No	PhD student or above	Yes	Yes	−0.12	94
Ihmels, Shemla, & Wegge, 2019b	3	Gender	Neutral/unclear	Female Male	48.19	No	.071	No	No	PhD student or above	Yes	Yes	−0.09 0.04	96 95
Kelly, Brown, Diekman, & Schneider, 2018	2	Race/Ethnicity	NA	NA	44.32	No	NA	No	Yes	PhD student or above	No	No	1.30	88
Kleineberg, 2018	2	Gender	Neutral/unclear	Unknown	64.89	No	.072	No	No	Student	Yes	Yes	0.39	423
Kührt & Hilse, 2012		Gender	Neutral/unclear	Unknown	46.19	No	.085	No	No	Student	Yes	Yes	0.33	194

Table 2 (*continued*)

Report	Study	Target characteristic	Type of crisis	Gender of previous leader(s)	Participant gender (%female)	Undergraduate sample	GII	Published	Original authors	Researcher experience	Manipulation check	DV reliable	*g*	*N*
Kulich, Iacoviello, & Lorenzi-Cioldi, 2018		Gender	Neutral/unclear	Male	50.00	No	.04	No	Yes	PhD student or above	Yes	Yes	−0.22	110
Kulich, Ryan, & Haslam, 2012		Race/Ethnicity	NA	NA	48.60	No	NA	Yes	No	PhD student or above	Yes	Yes	−0.04	108/113
Kulich, Ryan, & Haslam, 2006a		Race/Ethnicity	NA	NA	58.00	No	NA	Yes	No	PhD student or above	Yes	Yes	0.15	50
MacLean, 2015		Gender	Feminine	Male	46.96	No	.206	Yes	No	Student	Yes	Yes	0.17	526
McCarty & Kelly, 2011		Gender	Feminine	Unknown	28.74	Yes	.237	No	No	PhD student or above	Yes	Yes	0.06	87
Morgenroth, 2012	2	Gender	Neutral/unclear	Female Male	50.43	Yes	.085	No	No	Student	No	No	−0.26 −0.34	66 51
Poppe & Wuttig, 2011		Race/Ethnicity	NA	NA	70.00	Yes	NA	No	No	Student	Yes	Yes	0.28	79
Rink, Ryan, & Stoker, 2013	2	Gender	Neutral/unclear	Unknown	44.89	No	.047	Yes	Yes	PhD student or above	Yes	Yes	0.19	176
Ryan & Haslam (2004a)		Race/Ethnicity	NA	NA	61.76	Yes	NA	Yes	No	PhD student or above	No	Yes	0.07	36
Ryan & Haslam (2004b)		Race/Ethnicity	NA	NA	40.21	Yes	NA	Yes	No	PhD student or above	No	Yes	−0.21	94
Ryan & Haslam (2004c)		Gender	Neutral/unclear	Unknown	76.06	Yes	.206	Yes	No	PhD student or above	No	Yes	0.20	71
Ryan & Haslam (2007)		Gender	Feminine	Unknown	84.30	Yes	.206	Yes	No	PhD student or above	No	No	0.11	121
Ryan, Haslam, & Kulich, 2010	2	Gender	Neutral/unclear	Unknown	30.00	Yes	.172	Yes	Yes	PhD student or above	Yes	Yes/No[a]	−0.02	79/80
Schürmann, 2017		Gender	Feminine Masculine	Male	63.19	No	.072	No	No	Student	Yes	Yes	−0.00 −0.22	120 117

Note. *g* = Hedges' *g* (bias-corrected) where positive values indicate higher ratings of women or minorities as suitable leaders in crisis situations; GII = gender inequality index. Study number is only indicated for reports that contain multiple studies. Type of crisis, gender of previous leader(s), and GII are only provided for studies in which target gender was manipulated. Multiple *N*s are displayed when sample sizes varied for different measures.

[a] This study contained two dependent variables, one of which was reliable.

Table 3
Overview of Reports Included in the Leadership Selection Meta-Analysis

Report	Study	Target characteristic	Type of crisis	Gender of previous leader(s)	Participant gender (%female)	Undergraduate sample	GII	Published	Original authors	Researcher experience	Manipulation check	*OR*	*N*
Aelenei, Assilamehou-Kunz, Iacoviello, & Kulich, 2019		Race/Ethnicity	NA	NA	67.65	No	NA	No	No	Student	No	2.07	132
Ashby et al., 2007		Gender	Feminine	Unknown	66.67	Yes	.206	Yes	Yes	Student	Yes	2.46	111
Bain et al., 2011		Gender	Neutral/unclear	Unknown	65.38	Yes	.129	No	No	PhD student or above	Yes	1.33	130
Bain & Bongiorno, 2011		Gender	Neutral/unclear	Unknown	31.58	Yes	.129	No	No	PhD student or above	Yes	0.71	95
Belén, 2018		Gender	Neutral/unclear	Unknown	50.00	No	.080	No	No	Student	No	1.56	100
Boonsom, 2013		Gender	Neutral/unclear	Female	52.00	No	.317	No	No	Student	No	1.09	100
				Male								1.81	100
Bruckmüller et al., 2008		Gender	Neutral/unclear	Unknown	62.96	Yes	.092	No	No	Student	Yes	3.75	52
Bruckmüller & Branscombe, 2010	1	Gender	Neutral/unclear	Female	47.06	Yes	.242	Yes	No	PhD student or above	Yes	1.14	66
				Male								3.70	53
Gazijeva, 2017		Gender	Manipulated	Unknown	42.23	No	—	No	No	Student	Yes	100.46[a]	204
Haslam & Ryan, 2008	1	Gender	Neutral/unclear	Unknown	65.59	Yes	.206	Yes	Yes	PhD student or above	No	3.93	87
	2	Gender	Feminine	Unknown	71.76	Yes	.206	Yes	Yes	PhD student or above	No	6.18	83
Hunt-Earle, 2012		Gender	Neutral/unclear	Unknown	50.00	No	.172	Yes	No	PhD student or above	No	1.83	40
Ihmels et al., 2017	1	Gender	Neutral/unclear	Female	45.31	Yes	.072	No	No	PhD student or above	Yes	0.42	96
				Male								1.18	96
	2	Gender	Neutral/unclear	Unknown	59.55	Yes	.072	No	No	PhD student or above	Yes	0.71	172
Ihmels et al., 2019b	3	Gender	Neutral/unclear	Female	48.19	Yes	.071	No	No	PhD student or above	Yes	0.61	97
				Male								0.81	96
Kulich, Aelenei, Assilaméhou-Kunz, & Iacoviello, 2019		Race/Ethnicity	NA	NA	85.71	No	NA	No	No	PhD student or above	No	0.90	182
Kulich, Gartzia, & Komarraju, 2019		Gender	Feminine	Unknown	46.60	Yes	.087	No	No	PhD student or above	No	1.79	193
			Masculine									1.32	191
Kulich et al., 2018		Gender	Neutral/unclear	Unknown	50.00	No	.04	Yes	No	PhD student or above	Yes	1.33	110
Kulich, Lorenzi-Cioldi, Iacoviello, Faniko, & Ryan, 2015	1[b]	Gender	Neutral/unclear	Unknown	67.09	Yes	.048	Yes	Yes	PhD student or above	Yes	0.42	69
												2.84	70
	2	Gender	Neutral/unclear	Unknown	59.14	Yes	.048	Yes	Yes	PhD student or above	Yes	2.69	93
Kulich et al., 2006a		Race/Ethnicity	NA	NA	58.33	No	NA	No	Yes	PhD student or above	Yes	1.00	40
Kulich, Ryan, & Haslam, 2006b		Race/Ethnicity	NA	NA	61.43	No	NA	No	Yes	PhD student or above	Yes	0.82	66
MacLean, 2015		Gender	Feminine	Male	46.96	No	.206	No	Yes	Student	Yes	1.73	483
McCarty & Kelly, 2011		Gender	Feminine	Unknown	28.74	Yes	.237	No	No	PhD student or above	Yes	1.73	83
Morgenroth, 2012	2	Gender	Neutral/unclear	Female	50.43	Yes	.085	No	No	Student	No	1.05	66
				Male								0.67	51
Nadler & Bailey, 2015		Gender	Neutral/unclear	Unknown	76.96	Yes	.212	Yes	No	PhD student or above	No	1.58	76

Table 3 (*continued*)

Report	Study	Target characteristic	Type of crisis	Gender of previous leader(s)	Participant gender (%female)	Undergraduate sample	GII	Published	Original authors	Researcher experience	Manipulation check	*OR*	*N*
Ryan & Haslam (2004a)		Race/Ethnicity	NA	NA	61.76	Yes	NA	No	Yes	PhD student or above	No	2.25	35
Ryan & Haslam (2004b)		Race/Ethnicity	NA	NA	40.21	Yes	NA	No	Yes	PhD student or above	No	0.92	87
Ryan & Haslam (2004c)		Gender	Neutral/unclear	Unknown	76.06	Yes	.206	No	Yes	PhD student or above	No	2.14	71
Ryan et al., 2010	2	Gender	Neutral/unclear	Unknown	30.00	Yes	.172	Yes	Yes	PhD student or above	Yes	6.26	73
Schürmann, 2017		Gender	Feminine Masculine	Male	63.19	No	.072	No	No	Student	Yes	1.09 0.65	102 101
Verbeek, 2015		Gender	Neutral/unclear	Unknown	56.17	No	.045	No	No	Student	No	0.59	190
Wilson, 2010		Gender	Neutral/unclear	Unknown	47.96	No	.242	No	No	PhD student or above	Yes	1.97	269

Note. *OR* = odds ratio, where values above 1 indicate increased likelihood of selecting women or minorities for crisis situations; GII = gender inequality index. Study number is only indicated for reports that contain multiple studies. Type of crisis, gender of previous leader(s), and GII are only provided for studies in which target gender was manipulated.
[a] This outlier was excluded from the analyses. [b] This study had two crisis and control conditions. We therefore report two effect sizes.

differences (e.g., the average share price in the previous year for companies in which women vs. men had been appointed).[2] When multiple measures were reported for the predictor or the outcome, we coded all of them and combined them in the analyses, with some exceptions: When studies included data indicating a *change* in number of women or members of underrepresented ethnic and racial groups as well as static numbers, we only included the change data and when a study reported both change in performance and static performance, we included the change in performance data (see above).

Given that we did not restrict the meta-analysis to a specific domain (e.g., management), the measures of performance varied (see Table 1). We coded the domain (e.g., management, politics, education) to capture some of this heterogeneity in measures of performance (see Table 5 for a summary of all the moderators that we coded and whether they were intended as control variables or intended to answer theoretical questions). In the management domain, we further coded stock-based as compared with accountancy-based measures of performance, but other domains did not have sufficient variability in the measurement of performance. For example, although there were several effect sizes for the political domain, these studies almost exclusively measured performance based on winnability of the seat (based on the party's performance in the previous election).

We also coded which underrepresented group was the target of the glass cliff (women vs. underrepresented racial and ethnic groups), as well as a measure of gender inequality in the country in which the study was conducted. The latter was only coded for studies examining gender, using the gender inequality index (GII) provided by the UN, which measures gender inequalities in three aspects of human development: reproductive health, empowerment, and economic status (see United Nations Development Programme, 2019). We coded the GII for the relevant country and year of appointment. When the data spanned multiple years, we averaged the GII across the years. When data were not available for the exact year, we approximated the value from available data (e.g., if data were only available for 2005 and 2010, we would use the value for 2010 when we needed data for 2009 but the average between 2005 and 2010 when we needed data for 2008). Moreover, we coded whether the report was published, whether one of the original authors was involved in the study, and the study quality variables described above.

Coding of experimental studies. For experimental studies, we coded measures of leadership suitability/ability and selection measures (i.e., whether a man or woman was chosen for the leadership positions). In a typical study, participants read about an organization that was either doing well or doing poorly (manipulated between participants in all but one study) and looking for a new leader. They were then presented with one or more potential leaders and asked to evaluate them. In studies in which target characteristic (gender or race/ethnicity) was manipulated between participants, they were presented with one potential leader and asked to evaluate this target in terms of leadership ability or suitability. We coded the levels of leadership ability/suitability of the different candidates in the different performance conditions. When multiple measures of leadership ability/suitability were provided, we computed an effect size for each. We then combined

Table 4
List of Excluded Reports and Studies

Report	Reason for exclusion
Adams, Gupta, & Leeth, 2009	Overlapping data
Andersen, 2013	Overlapping data
Bechtoldt, Voigt, & Bannier, 2016	Overlapping data
Bechtoldt, Bannier, & Rock, 2019, Study 1	Overlapping data
Brady, Isaacs, Reeves, Burroway, & Reynolds, 2011	Overlapping data
Bruckmüller, 2007	Overlapping data
Brunner, 2014	Missing information
Burton & Grappendorf, 2015	Missing information
Carroll, Hennessey, & MacDonald, 2013	Overlapping data
Çelebi & Saatci, 2016	Missing information
Cook & Glass, 2014a	Overlapping data
Cook & Glass, 2014c	Overlapping data
Del Prete & Stefani, 2015	Missing information
Dobbin & Jung, 2011	Overlapping data
Elsaid & Ursel, 2018	Overlapping data
Gabarrot, Bry, De Oliveira, & Dietz, 2014	Missing information
Gündemir et al., 2019, Study 1	Overlapping data
Kurt, 2011	Overlapping data
MacDonald, 2011	Overlapping data
Morgenroth, 2012, Study 1	Overlapping data
Mulcahy & Linehan, 2014	Overlapping data
O'Brien, 2015	Missing information
Oelbaum, 2016, Study 1	Missing information
Ryan et al., 2010	Overlapping data
Stokman, 2011	Overlapping data
Thomas & Bodet, 2013	Missing information
Uyar, 2011	Overlapping data

Table 5
Moderators Included in Each of the Three Meta-Analyses

Analysis	Archival	Leader suitability	Leader selection
Full gender analysis			
Domain	*		
History of leadership		*	*
Type of crisis		*	*
Gender inequality index of country	*	*	*
Participant gender		*	*
Undergraduate sample		•	•
Research group	•	•	•
Type of appointment measure	•		
Type of performance measure	•		
Time between performance and appointment	•		
Researcher experience	•	•	•
Manipulation check		•	•
Reliability		•	
Race subanalysis			
Gender versus race	*	*	*
Femininity of race stereotypes		*	*
Gender subanalysis			
Accountancy versus stock-based	•		

• Indicates a moderator that was included as a control variable (e.g., study quality indicators or methodological details). * Indicates a moderator that was included for theoretical reasons.

[2] Two studies (Bowles, 2013; Robinson, 2019) reported a mix of continuous and dichotomous predictor variables. In these cases, we only used the continuous variables.

them in the analyses (details in in the Analytic Strategy section) because the measures were conceptually similar (e.g., "This candidate would be good leader" vs. "This candidate will bring required skills to the job"), and we had no theoretical reason to expect a difference. For studies including both measures, the average correlation was $\bar{r} = .57$.

In studies in which target characteristic was a within-participants factor, they responded to similar leadership suitability/ability measures for multiple participants. In addition to these measures, participants often ranked the candidates as well, which we term *leadership selection*. In these cases, we coded how many participants ranked the male/white candidate first in the different performance conditions and how many participants ranked the female/minority candidate first in the two performance conditions. We examined these leadership selection measures in a separate meta-analysis from leadership suitability evaluations because evaluations of women's leadership and their subsequent selection (or not) for a leadership position do not always map onto one another.

When a study reported multiple crisis conditions that were not clearly related to femininity or masculinity (e.g., severe crisis and moderate crisis) or multiple minority group conditions (e.g., African Americans and Asian Americans), we collapsed numbers across these conditions,[3] including cases in which the glass cliff was predicted for one of the conditions, but not the other (e.g., when the crisis was described as severe or moderate). This decision resulted in a conservative estimate of effect sizes, but we chose this strategy because there were relatively few studies testing the same moderators or providing sufficient information to code for the moderators.

To examine evidence regarding the explanations for the glass cliff, we coded whether the previous leaders were male, female, or whether this information was unknown, as well as whether the specific crisis called for stereotypically feminine skills, masculine skills, or a set of skills that was not clearly gendered. More specifically, coders received the following instructions:

> Is the crisis one that likely requires stereotypically feminine or masculine attributes? For example, a technical failure could be seen as masculine (e.g., requiring technical expertise), while a scandal or lack of popularity could be seen as feminine (e.g., requiring social skills). Other crises might be fairly neutral, or it is unclear whether they require feminine or masculine skills.

When either of these variables were manipulated between participants in the study, we coded values separately for the two conditions. We also coded the GII for the year and country in which the data were collected. When the year of data collection was not known, we estimated that data were collected two years prior to the publication of a study and in the same year for BSc and MSc theses. These three variables were only coded for studies in which the gender (rather than race or ethnicity) of the potential leader was manipulated.

In addition, we coded target characteristic (gender or race/ethnicity), percentage of female and male participants,[4] whether participants were undergraduate students, whether the study was published, whether one of the original authors was involved in the study, whether it was a student project, whether the dependent variable was reliable (consisting of multiple items with an $\alpha \geq .70$), and whether a manipulation check indicated that performance had been successfully manipulated.

Coding reliability. We calculated Cohen's kappa for categorical variables and Intraclass Correlation Coefficients (ICCs) for continuous variables to assess intercoder reliability.[5] All ICCs were above .80 (ranging from .81 to >.99), indicating good reliability, and 80% of all ICCs were above .90, indicating excellent reliability (Koo & Li, 2016). Cohen's κ values ranged from 0.32, indicating fair levels of agreement, to 1.00, indicating complete agreement (McHugh, 2012). Although the majority of κs indicated substantial (between .61 and .80; 23.08% of all variables) or almost perfect (>.80; 69.23% of all variables) agreement, the intercoder reliability was lower for two variables: For the archival studies, the variable "time" (i.e., whether the performance variable preceded the appointment variable by no more than a year) yielded a $\kappa = .55$. Moreover, for the studies included in the leader evaluation meta-analysis, the reliability for "type of crisis" (i.e., whether a crisis required feminine or masculine attributes) was $\kappa = .32$. As a result, the coding for these two variables was discussed extensively by the coders until a consensus was reached. All other conflicts were also resolved through discussion.

Analytic Strategy

We ran three separate sets of meta-analyses: (a) archival studies (Hedges' *g*), (b) experimental studies with leadership rating scales (Hedges' *g*), and (c) experimental studies with binary leadership selection measures (odds ratios reflecting choice of a member of an overrepresented group over a member of an underrepresented group). We will address our analytic strategy for each of these in turn in the sections below. We calculated all effect sizes so that positive values reflected preference for members of underrepresented groups in crisis relative to no-crisis conditions. We first ran all analyses on the effect sizes comparing men with women, only adding the effect sizes for members of underrepresented racial and ethnic groups as part of a separate moderator analysis.

For all meta-analyses, to estimate the overall mean weighted effect sizes, we ran random effects models with the MeanES.sps SPSS macro developed by Wilson (2005), using method of moments. We opted for random effects models to make generalizations beyond the included effect sizes and because we expected the variability in methods and samples to lead to systematic variability across studies. Our random effects models weighted effect sizes using the inverse variance of each effect and the between-studies variance component.

[3] An exception to this rule was one study in which two between-participants crisis conditions were matched to two separate control conditions (Kulich et al., 2015, Study 1). Here, we did not collapse across conditions but coded them as two separate effects.

[4] When samples were split by condition based on type of crisis or gender of the previous leaders, the numbers of female and male participants in this subsample were not always available. When they were not, we applied the proportion of the entire sample to the subsample.

[5] Please note that, as one of the two original coders left academia before this project was completed, a total of four coders were involved in this project (the first coder was the same throughout). This violates the assumptions of the calculations for the ICC and Cohen's kappa. However, we judged that these calculations were still more appropriate than the alternative—assuming that two randomly selected coders from a large pool of coders coded each variable, particularly given that the majority of coding for each variable was done by the same two coders.

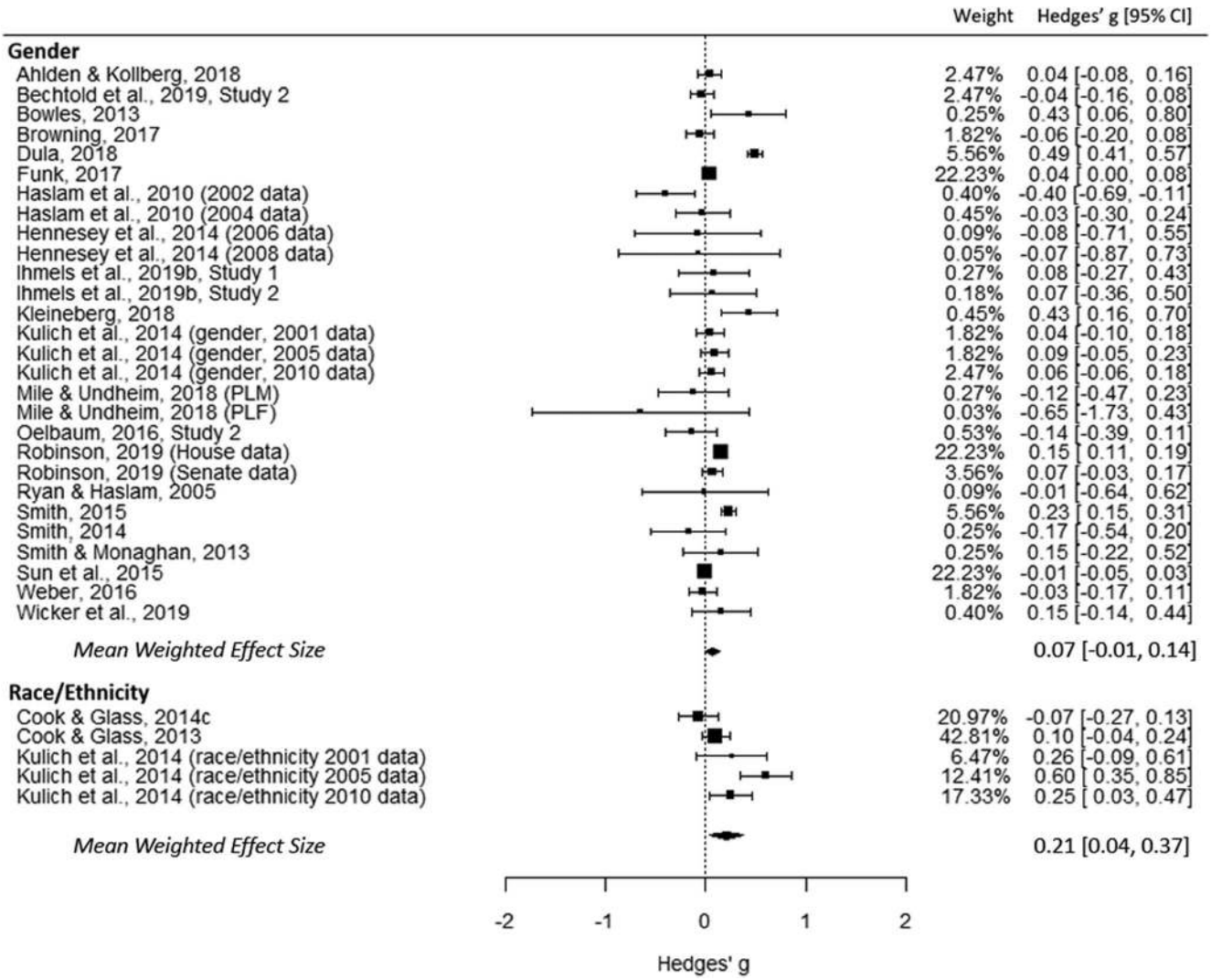

Figure 2. Forest plot of archival Hedges' *g* effect sizes with 95% confidence intervals. Effect sizes above 0 indicate a glass cliff effect (higher appointment of women or minorities in crisis situations). Percentages, as well as the size of each point, reflect the relative weight of the effect. PLF (previous leader female) and PLM (previous leader male) designate independent samples within a study where the previous leader was either female or male respectively.

To assess the heterogeneity of effect sizes, we reported Cochrane's Q tests (Q_W), which tests whether the between-study variance component (τ^2) is significantly different from zero. We also computed I^2, which indicates the percentage of heterogeneity stemming from systematic differences across studies, rather than from sampling error. Finally, we reported T, which describes the random effects variance component.

We ran categorical moderation analyses using the MetaF.sps SPSS macro using mixed effects models with maximum likelihood estimation. As recommended by both Lipsey and Wilson (2001) and Borenstein, Hedges, Higgins, and Rothstein (2009), we tested between-study moderation with fixed effects (reflecting systematic study-level variability), but retained the random effects component that also assumes remaining study-level variability in addition to the participant-level sampling error. We used a common value for the variances of each level of the moderator, as recommended when there are small numbers of effects in each group. Although mixed effects models can be underpowered for testing moderation (Lipsey & Wilson, 2001), we nonetheless opted for this approach because the assumption of homogeneity in fixed effects models was not appropriate for these studies. In these analyses, we reported Q_B, which tests whether there is a significant amount of heterogeneity that can be attributed to differences between moderator subgroups.

We ran continuous moderation analyses using the MetaReg.sps SPSS macro using mixed effects models with maximum likelihood estimation. We tested moderation with fixed effects (reflecting systematic study-level variability) but retained the random effects component that also assumes remaining study-level variability in addition to the participant-level sampling error (i.e., a random

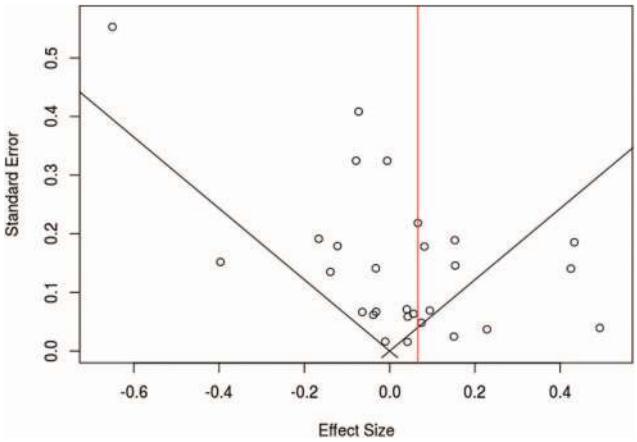

Figure 3. Funnel plot of archival Hedges' *g* effect sizes with black contour lines added at the *p* value cutpoint (.05). Effect sizes located outside of the contour lines represent statistically significant effects, while those within the contour lines are not statistically significant. Effect sizes above 0 indicate a glass cliff effect (higher appointment of women or minorities in crisis situations). The vertical line indicates the meta-analytic mean estimate. See the online article for the color version of this figure.

intercept and fixed slopes model). In these analyses, we reported Q_R, which tests whether the moderator accounts for significant variability in the model. To visualize the effect of continuous moderators on the glass cliff, we generated meta-regression plots with confidence bands using the moving constant technique (Johnson & Huedo-Medina, 2011) in Stata. After running bivariate analyses, we ran a multiple meta-regression model that included all moderators that were statistically significant in the bivariate analyses in order to maximize statistical power. Categorical variables were dummy-coded.

Archival Studies

For studies that reported means and standard deviations, we calculated Hedges' *g* (a variant of Cohen's *d*) reflecting poorer performance before appointments of members of underrepresented groups, relative to appointments of members of overrepresented groups, using the weighted pooled standard deviation. For a subset of studies that instead reported point-biserial correlations, we converted the effect sizes to an estimate of Hedges' *g* using the formula recommended by Lipsey and Wilson (2001; see the Appendix for formulae).

If there was more than one relevant dependent measure for a single study (e.g., change in Tobin's Q and change in share price), we created a mean of the effect sizes and calculated the variance based on the formula recommended by Borenstein and colleagues (2009), as preregistered (https://osf.io/b8tzq/?view_only=d0c1b3b2fb144d979f01bfc6a3ea39a7).[6] In cases when there were missing data for one measure, we used the measure with more data points. Finally, we adjusted Hedges' *g* effect sizes to correct for bias (Hedges, 1984) and weighted effect sizes using the inverse variance and the between-studies variance component (Lipsey & Wilson, 2000).

Leadership Suitability Ratings

Combining study designs. We first tested whether it was reasonable to combine within and between-participants studies by testing study design as a moderator of the leadership rating effect sizes. We tested this moderator because different designs should only be combined when they measure the same treatment effect (Morris & Deshon, 2002), and the effect sizes must use identical calculations regardless of design.

Effect size calculations. For all studies, we obtained means and standard deviations of the leadership suitability ratings of members of overrepresented and underrepresented groups separated by crisis and no-crisis condition. Because calculating the magnitude of the glass cliff required taking into account both crisis condition and gender of the person being rated, we calculated effect sizes in line with Morris and DeShon's (2002) recommendations for 2 × 2 designs comparing subgroups (Morris & DeShon, 2002, p. 114, Table 1).

We first calculated a Hedges' *g* effect size reflecting preference for members of underrepresented groups over members of overrepresented groups, separately for both the crisis and no-crisis conditions. Specifically, we subtracted leadership ratings for members of overrepresented groups from ratings for members of underrepresented groups and divided the difference score by the weighted pooled standard deviation.[7] This calculation is standard for between-participants studies, and we applied this formula to within-participants studies as well to ensure comparable effect sizes across design (because the research is focused on group differences, rather than individual-level change; Morris & DeShon, 2002).

We then translated these two Hedges' *g* effect sizes into a single effect size by subtracting the no-crisis condition from the crisis condition effect size (Morris & DeShon, 2002; Borenstein et al., 2009)—all formulae for these calculations appear in the Appendix. Finally, we corrected the effect sizes for bias and weighted effect sizes using the inverse variance and the between-studies variance component.

Sampling variance calculations. To calculate the sampling variance for the 22 effect sizes where target gender or race/ethnicity was within-participants, we used Morris and DeShon's (2002, Table 2) formula for pretest-posttest designs in a raw score metric. We

[6] Haslam and colleagues (2010) argued and found evidence that the glass cliff is more likely to apply to subjective indicators of company performance (e.g., stock-based measures or other performance indicators where social factors, such as the evaluation by others, play a role) than objective indicators (e.g., accountancy-based measures that reflect actual company performance). Thus, for studies in the management domain, we first tested whether the magnitude of the glass cliff differed for stock-based and accountancy-based measures, in line with our preregistration. Inconsistent with expectations, the magnitude of the stock-based effect size ($d = .04$, $p = .39$) was comparable with that of the accountancy-based effect size ($d = -.02$, $p = .25$), $Q_B(1) = 1.36$, $p = .24$. We next collapsed these measures in the meta-analysis using the same approach as for other studies including multiple measures.

[7] For one study (Bruckmüller, 2007), the standard deviation for one condition was missing (female candidate, crisis condition), so we used the standard deviation for the male candidate crisis condition in the effect size calculation, rather than the pooled standard deviation.

Table 6
Summary of Archival Moderation Analyses

Moderator	*g*	95% CI	*p*	*k*	*Q*	*p*	I^2	*T*
Domain					103.00	<.001		0.03
Management	−.02	[−.07, .02]	.32	14	9.90	.70	0.00	
Politics	.08	[.04, .11]	<.001	10	16.67	.05	46.01	
Education	.24	[.15, .33]	<.001	2	—	—	—	
Sports	.15	[−.14, .44]	.30	1	—	—	—	
Nonprofit	.49	[.40, .59]	<.001	1	—	—	—	
Research group					1.44	.23		0.14
Original	−.01	[−.16, .13]	.86	6	4.18	.52	0.00	
Other	.09	[.01, .17]	.02	22	22.06	.40	4.81	
Minority group					2.36	.12		0.15
Women	.07	[−.01, .14]	.07	28	25.44	.55	0.00	
Racial/Ethnic	.21	[.04, .37]	.01	5	6.53	.16	38.74	
Appointment measure					1.31	.25		0.14
Presence of female leaders	.13	[.01, .25]	.04	8	18.67	.01	62.51	
Appointment of female leaders	.04	[−.04, .12]	.29	20	9.06	.97	0.00	
Performance measure					1.81	.18		0.14
Static performance	.09	[.02, .17]	.02	19	26.23	.09	31.38	
Change of performance	−.01	[−.14, .12]	.88	9	0.90	>.99	0.00	
Time					0.94	.33		0.15
More than one year	.10	[.001, .21]	.05	9	4.88	.77	0.00	
Year or less	.04	[−.06, .13]	.46	19	21.09	.27	14.65	
Researcher experience					0.04	.84		0.15
Student	.06	[−.07, .18]	.37	8	6.74	.46	0.00	
PhD student or above	.07	[−.01, .16]	.10	20	20.07	.39	5.33	
Continuous moderators (*B*)					0.02	.90		
Gender inequality index	.05	[−.73, .83]	.90	28	26.76	.42		0.15

Note. *g* is the mean Hedge's *g* effect size (bias-corrected) where positive values indicate higher appointment of women or minorities in crisis situations. For the continuous moderator analysis, the unstandardized *B* coefficient is reported instead, and it describes the relationship between the gender inequality index of the country under consideration and the presence of a glass cliff. CI = confidence interval. *k* indicates the number of effect sizes included in each analysis. Q_B tests for differences across subgroups and is reported under *Q* in the first row for each moderator. Q_W tests homogeneity within subgroups and is reported under *Q* in that subgroup's row. Q_R tests whether the moderator accounts for significant variability in the model and is reported under *Q* in the case of continuous moderation analyses. I^2 indicates the percentage of heterogeneity stemming from systematic differences across studies, rather than from sampling error. *T* describes the standard deviation of the true effect sizes across studies. Dashes indicate insufficient degrees of freedom to produce an estimate.

implemented this formula[8] separately for the crisis and no-crisis conditions and then summed the variances (Borenstein et al., 2009, p. 223).

To calculate the sampling variance for the 14 effect sizes where target gender or race/ethnicity was between-participants, we used the formula recommended for independent groups posttest designs in raw score metric (Morris & DeShon, 2002, Table 2). We computed the variance for the crisis and no-crisis conditions separately and then added them together (Morris & DeShon, 2002).

Table 7
Summary of Archival Multiple Meta-Regression Analysis

Moderator	*B*	95% CI	β	*p*
Domain				
Politics	.11	[.02, .20]	.33	.01
Sports	.19	[−.13, .51]	.14	.24
Nonprofit and education	.40	[.28, .52]	.88	<.001

Note. The multiple meta-regression analysis included all moderators for which $p < .05$ in bivariate analyses. *B* = unstandardized regression coefficient; CI = confidence interval; β = standardized beta coefficient. $k = 28$ and indicates the number of effect sizes included. The domain variable was dummy coded such that management was the reference group (0), and other domains were coded as 1.

If there was more than one relevant dependent measure for a single study (e.g., leadership suitability and ability), we created a mean of those effect sizes and calculated the variance based on the formula recommended by Borenstein and colleagues (2009; see the Appendix).

Leadership selection. Using cell frequencies, we calculated odds ratios that reflected the odds of selecting a member of an underrepresented group relative to a member of an overrepresented group for leadership in crisis relative to no-crisis situations. We conducted the meta-analysis on the log odds ratios, weighting effect sizes using the inverse variance and the between-studies variance component of the log odds ratios (Lipsey & Wilson, 2001). The aggregate effect size was then converted back to an odds ratio.

Publication bias. For each meta-analysis, we checked for evidence of publication bias using Vevea and Hedges' weight-function model application (Hedges & Vevea, 1996; Vevea & Hedges, 1995;

[8] The population effect size (Hedges' *g*) needed for this formula was estimated using the unweighted average effect size of gender (Morris & DeShon, 2002). The population correlation (ρ) between male and female ratings was estimated by meta-analyzing the available correlations. Because there was significant heterogeneity, we opted to instead use each study's respective correlation in the formula, but we used the population correlation for studies where we were unable to obtain the correlation.

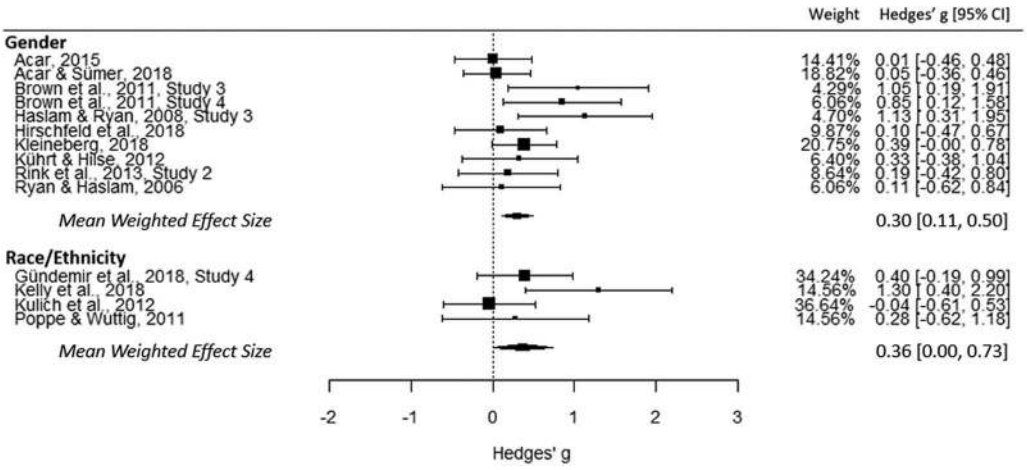

Figure 4. Forest plot of experimental Hedges' *g* effect sizes (leadership suitability ratings) for between-participants studies with 95% confidence intervals. Effect sizes above 0 indicate a glass cliff effect (higher ratings of women or minorities as suitable leaders in crisis situations). Percentages, as well as the size of each point, reflect the relative weight of the effect.

https://vevealab.shinyapps.io/WeightFunctionModel/). We used a selection model with heterogeneous effect sizes that assumes that both statistically significant and nonsignificant effect sizes may be published (but with different likelihoods), as recently recommended by multiple simulation studies (McShane, Böckenholt, & Hansen, 2016; also see Carter, Schönbrodt, Gervais, & Hilgard, 2018, for an overview of the shortcomings of other advanced techniques with typical psychological data). This method generates an estimate of the bias-corrected effect size and compares the fit of the bias-adjusted model to the unadjusted original model. Because the meta-analyses contained relatively few effect sizes, selection was based on only two *p* value intervals: .0 to .05 and .05 to 1.00. We chose $p = .05$ as the cutoff because Hedges and Vevea (1996) recommend setting critical alpha levels as cutoff points when these are psychologically meaningful in a discipline. The application also generates funnel plots with contour lines at *p* value cutpoints (see Figures 3, 5, 7, and 10).

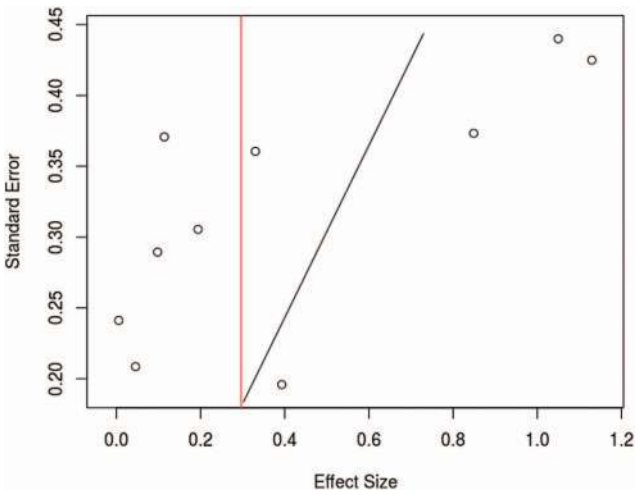

Figure 5. Funnel plot of experimental Hedges' *g* effect sizes (leadership suitability ratings) for between-participants studies with a black contour line added at the *p* value cutpoint (.10). We used a *p* value cutpoint of .10 because the model did not converge at *p* values of .05 or .01. Effect sizes located outside of the contour line (i.e., to the right) represent statistically significant effects, whereas those within the contour line (i.e., to the left) are not statistically significant. Effect sizes above 0 indicate a glass cliff effect (higher ratings of women or minorities as suitable leaders in crisis situations). The vertical line indicates the meta-analytic mean estimate. See the online article for the color version of this figure.

Results

Description of Studies in the Meta-Analyses

We coded a total of 74 independent studies conducted between 2004 and 2019. The time periods examined by the archival studies included the years 1983 to 2016. The majority of experimental studies were conducted in the United Kingdom (24.00%), United States (24.00%), and Germany (20.00%). The archival studies primarily examined data from the United States (45.83% of studies) and the United Kingdom (16.67% of studies). Gender inequality was generally low, but still varied between countries and between years in the same countries.

The majority of included studies examined the glass cliff in the context of gender. However, there were three archival studies, eight leader evaluation studies, and six leader selection studies examining the glass cliff in the context of race or ethnicity. More information on the coded information for all studies included in the three meta-analyses can be found in Tables 1 through 3.

Archival Studies

Summary analyses. The weighted aggregate effect size across 28 independent samples (only including studies focusing on leaders' gender) showed that women were indeed more likely to be

Table 8
Summary of Leadership Suitability Moderation Analyses for Between-Participants Studies

Moderator	*g*	95% CI	*p*	*k*	*Q*	*p*	I^2	*T*
Research group					0.27	.61		0.18
Original	.39	[−.01, .79]	.06	3	4.00	.14	50.00	
Other	.27	[.07, .47]	.007	7	8.68	.19	30.88	
Minority group					0.08	.77		0.11
Women	.30	[.11, .50]	.002	10	11.68	.23	22.95	
Racial/Ethnic	.36	[.002, .73]	.05	4	5.73	.13	47.64	
Sample					0.11	.74		0.07
Undergraduate	.35	[.01, .68]	.04	4	6.47	.09	53.63	
Other	.28	[.06, .50]	.01	6	5.80	.33	13.79	
Manipulation check					0.12	.73		0.00
Successful	.31	[.11, .52]	.003	7	7.39	.29	18.81	
Unsuccessful or absent	.24	[−.12, .60]	.19	3	5.44	.07	63.24	
Reliability					0.25	.61		0.06
High	.31	[.12, .50]	.001	9	12.27	—	34.80	
Low or not reported	.11	[−.62, .85]	.76	1	—	—	—	
Type of crisis					5.70	.06		0.00
Feminine crisis	.11	[−.61, .84]	.76	1	—	—	—	
Masculine crisis	.93	[.38, 1.49]	.001	2	0.12	—	—	
Neutral or unclear	.23	[.04, .42]	.02	7	7.12	.31	15.73	
Leadership history					—	—		—
Female leader	—	—	—	—	—	—	—	
Male leader	—	—	—	—	—	—	—	
Unknown	—	—	—	—	—	—	—	
Researcher experience					0.01	.92		0.00
Student	.31	[.02, .60]	.04	3	0.72	.70	0.00	
PhD student or above	.29	[.06, .51]	.01	7	12.22	.06	50.90	
Continuous moderators (*B*)								
Gender inequality index	−.57	[−2.09, 0.95]	.46	10	0.54	.46		0.00
Participant gender	.003	[−.01, .02]	.77	10	0.09	.77		0.00

Note. *g* is the mean Hedge's *g* effect size (bias-corrected) where positive values indicate higher ratings of women or minorities as suitable leaders in crisis situations. For the continuous moderator analysis, the unstandardized *B* coefficient is reported instead, and it describes the relationship between the moderator (gender inequality index of the country under consideration and the percentage of female participants) and the presence of a glass cliff. CI = confidence interval. *k* indicates the number of effect sizes included in each analysis. Q_B tests for differences across subgroups and is reported under *Q* in the first row for each moderator. Q_W tests homogeneity within subgroups and is reported under *Q* in that subgroup's row. Q_R tests whether the moderator accounts for significant variability in the model and is reported under *Q* in the case of continuous moderation analyses. I^2 indicates the percentage of heterogeneity stemming from systematic differences across studies, rather than from sampling error. *T* describes the standard deviation of the true effect sizes across studies. Dashes indicate insufficient degrees of freedom to produce an estimate.

appointed to leadership positions in times of crisis (see Figure 2), $g = 0.07$ ($T = 0.13$), 95% CI [0.001, 0.13], $p = .04$, although the effect size was relatively small.[9] There was a large amount of heterogeneity in the effect size, with 87.18% of the heterogeneity (I^2) stemming from systematic differences across studies rather than sampling error,[10] $Q(27) = 210.61$, $p < .001$.

Study quality. The majority of studies demonstrated high quality by (a) analyzing the selection or appointment of a female leader (66%) rather than the presence of female leaders, (b) reporting analyses in which the time between the performance indicator and the appointment of a female leader was a year or less (78%), and (c) being conducted by a researcher with more research experience (doctoral-level student or higher; 72%). Yet, only a minority of studies used a high-quality measure of performance (36%) by reporting change in performance over time rather than static measures.

Moderation analyses. Table 6 shows the statistical results for all moderation analyses. Out of nine potential moderators, the magnitude of the glass cliff effect only differed across organizational domain, $Q_B(4) = 103.00$, $p < .001$. When examined separately by domain, the effect held specifically for the education, political, and nonprofit domains, but not for the management or sports domains.

We next ran a multiple meta-regression analysis to compare the magnitude of the effect in each domain (see Table 7). We collapsed the education and nonprofit domains because of the small number of effect sizes and because educational institutions are usually nonprofit organizations. The domain variable was dummy coded with management as the reference group, which led to three variables in the regression: politics (1 = politics, 0 = other), sports (1 = sports, 0 = other), and nonprofit (1 = nonprofit, 0 = other). The analysis revealed that women were more likely to be appointed to precarious leadership positions in political and nonprofit

[9] In the initial analysis, there was one extreme outlier ($g = -1.40$) that was more than four standard deviations below the effect size mean. Upon closer inspection, this outlier comprised two separate effect sizes that had been aggregated to ensure independence of effects, one value of $d = -2.27$ and one of $d = -.69$. We excluded the extreme value of −2.27 and instead used the estimate of −0.69 to represent this study.

[10] Because of large variability in sample sizes, three effect sizes (Funk, 2017; Robinson, 2019; Sun, Zhu, & Ye, 2015) had more weight in the analysis than all of the other studies combined. A sensitivity analysis excluding these three studies did not substantially change the summary effect size ($g = .07$) or the results of the moderation analyses.

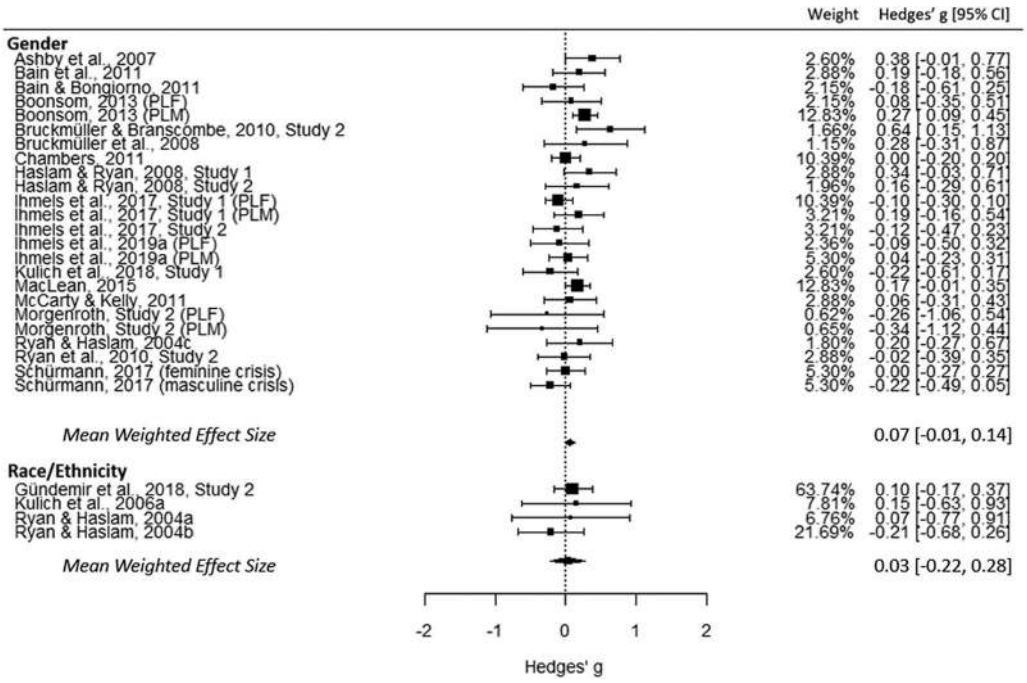

Figure 6. Forest plot of experimental Hedges' *g* effect sizes (leadership suitability ratings) for within-participants studies with 95% confidence intervals. Effect sizes above 0 indicate a glass cliff effect (higher ratings of women or minorities as suitable leaders in crisis situations). Percentages, as well as the size of each point, reflect the relative weight of the effect. PLF (previous leader female) and PLM (previous leader male) designate independent samples within a study where the previous leader was either female or male (manipulated between-participants), respectively.

roles than in management roles ($R^2 = .63$; $I^2 = .06$; $T = .06$). There was no difference between sports and management roles (see Table 7).[11] The estimated mean effect size when accounting for domain was $g = 0.09$.

Experimental Studies: Leadership Suitability Ratings

Between-participants analysis.

Summary analyses. The weighted aggregate effect size across 10 independent samples[12] (only including studies focusing on leaders' gender) showed that women were indeed rated as more suitable leaders in times of crisis (see Figure 4), $g = 0.32$ ($T = 0.19$), 95% CI [0.10, 0.54], $p = .004$. There was a low amount of heterogeneity in the effect size, with only 30.50% of the heterogeneity (I^2) stemming from systematic differences across studies rather than sampling error, $Q(9) = 12.95$, $p = .17$. Despite low and nonsignificant heterogeneity, we proceeded with mixed effect moderator analyses due to the small number of effect sizes (Borenstein et al., 2009).

Study quality. The majority of studies demonstrated high quality because they were (a) conducted by researchers with more experience (doctoral-level student or higher; 72.73%), (b) included a successful manipulation check (63.64%), and (c) reported reliability of .70 or greater for the dependent measure (81.82%).

[11] This finding was the same when including all control moderators (e.g., study quality indicators) in the same regression (see Table 5 for information about control moderators). However, two of the study quality variables became statistically significant in the meta-regression: The effects of student project and the time between measuring performance and the leadership appointment. Student projects showed a stronger glass cliff effect than nonstudent projects, $\beta = -0.21$, $p = .009$. Additionally, a stronger glass cliff emerged when there was less than a year between the measure of performance and the appointment (an indication of higher quality), $\beta = 0.87$, $p < .001$. Because these variables were not significant in bivariate analyses, it is unclear whether they are robust moderators.

[12] The experimental portion of the meta-analysis contained studies that manipulated candidate gender between ($k = 10$) and within-participants ($k = 17$). Although the methods used for between and within-participants studies were similar and there was no theoretical reason to expect a difference across study designs, we first compared the meta-analytic effects to determine whether it was appropriate to collapse across design type. Indeed, the magnitude of the effect differed by design type, $Q_B(1) = 5.10$, $p = .02$, and the glass cliff effect emerged only in between-participants studies, $g = 0.30$, $p = .002$, and not in within-participants studies, $g = 0.07$, $p = .11$. Given the difference across designs, it was not appropriate to meta-analytically combine the within and between-participants studies, so we proceeded with the full analyses separated by design type. Additionally, in a moderator analysis of all between and within-participants studies, only GII emerged as a significant moderator, $\beta = .51$, $p < .001$. When including GII, $\beta = .48$, $p = .001$, and study design, $\beta = -.27$, $p = .05$, together in a multiple meta-regression, they both remained statistically significant.

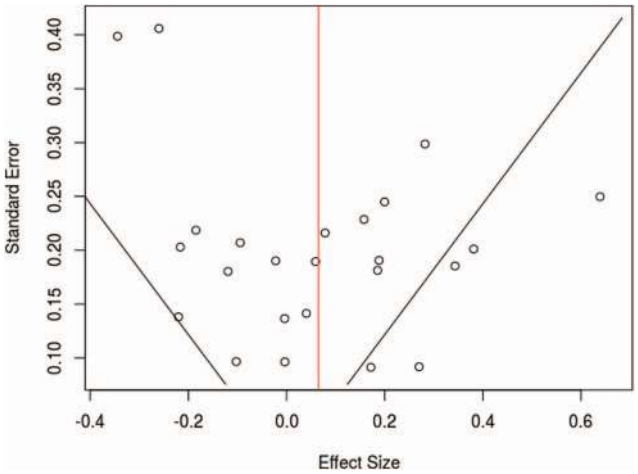

Figure 7. Funnel plot of experimental Hedges' *g* effect sizes (leadership suitability ratings) for within-participants studies with black contour lines added at the *p* value cutpoint (.05). Effect sizes located outside of the contour lines represent statistically significant effects, whereas those within the contour lines are not statistically significant. Effect sizes above 0 indicate a glass cliff effect (higher ratings of women or minorities as suitable leaders in crisis situations). The vertical line indicates the meta-analytic mean estimate. See the online article for the color version of this figure.

Moderation analyses. Of nine potential moderators, none were statistically significant in bivariate (see Table 8) or in multiple meta-regression analyses.

Within-participants analysis.

Summary analyses. The weighted aggregate effect size across 24 independent samples (only including studies focusing on leaders' gender) showed that women were not rated as more suitable leaders in times of crisis (see Figure 6), $g = 0.07$ ($T = 0.11$), 95% CI [−0.02, 0.15], $p = .12$.[13] There was a low amount of heterogeneity in the effect size, with only 30.47% of the heterogeneity (I^2) stemming from systematic differences across studies rather than sampling error, $Q(23) = 33.08$, $p = .08$. As with the between-participants analysis, we still examined moderator analyses in the subsequent sections and proceeded with mixed effects analyses.

Study quality. The majority of studies demonstrated high quality because they were (a) conducted by researchers with more experience (doctoral-level student or higher; 65.52%), (b) included a successful manipulation check (62.07%), and (c) reported reliability of .70 or greater in the dependent measure (75.86%).

Moderation analyses. Table 9 shows the statistical results for all moderation analyses. Out of nine potential moderators, only research group and gender inequality index of the country were statistically significant. Yet, in a multiple meta-regression including these two moderators ($R^2 = .54$), only the gender inequality index remained significant, showing that the glass cliff was larger in countries with more gender inequality (see Table 10).[14] Figure 8 visualizes this moderation pattern, suggesting that the glass cliff occurs in countries with a gender inequality index above approximately 0.15. This includes countries such as the United Kingdom, United States, and Thailand, but not Germany and Switzerland.

Race subanalysis. To explore evidence for or against the different explanations for the glass cliff, we also examined the magnitude of the effects across racial and ethnic groups. We collapsed across study design and did not interpret *p* values, because of the limited number of effect sizes. Our exploratory meta-analysis showed that the glass cliff effect was the strongest for Black, $g = 0.26$ ($k = 4$), and Middle Eastern, $g = 0.28$ ($k = 1$) leaders, as compared with East Asian, $g = 0.11$ ($k = 3$), and South Asian, $g = 0.07$ ($k = 1$) leaders. This pattern is inconsistent with the suggestion that underrepresented groups are appointed to leadership positions in times of crisis as a result of group stereotypes. Instead, it is more in line with the argument that organizations are motivated to signal change or that prejudice is driving the glass cliff. Because this analysis only included a small number of effects, results should be interpreted with caution.

Experimental Studies: Leader Selection

Summary analyses. The weighted aggregate effect size across 35 independent samples (only including studies focusing on the selection of female vs. male leaders) showed that women were indeed more likely to be selected for leadership positions in times of crisis (see Figure 9), $OR = 1.45$ ($T = 0.46$), 95% CI [1.17, 1.78], $p < .001$.[15] There was a moderate amount of heterogeneity in the effect size, with 57.47% of the heterogeneity (I^2) stemming from systematic differences across studies rather than sampling error, $Q(34) = 79.95$, $p < .001$.

Quality of studies. The majority of studies demonstrated high quality because they were conducted by researchers with more experience (doctoral-level student or higher; 68.57%) and included a successful manipulation check (62.86%).

Moderation analyses. Table 11 shows the statistical results for all moderation analyses. Of eight potential moderators, only gender of previous leader, research group, and gender inequality index of the country were statistically significant. Specifically, there was evidence for a glass cliff when participants had no information about the gender of previous leaders, but not when the previous leader was known to be either male or female. In addition, the glass cliff was larger in countries with more gender inequality and for studies conducted by the original authors.

We next ran a multiple meta-regression including all of the moderators that were significant in bivariate analyses (see Table 12).[16] The leadership history variable was dummy coded with unknown leader as the reference group, which led to two variables in the regression: one comparing effects with a history of male leadership to those with an unknown history of leadership, and one

[13] With the exception of one effect size, crisis condition was manipulated between participants. The effect size for the within-participants study was comparable to that of the between-participants study, so we included both study designs in the meta-analysis, $Q_B(1) = 0.28$, $p = .60$.

[14] This finding was the same when including all control moderators in the same regression.

[15] In the initial analysis, there was one extreme outlier ($OR = 100.46$; $ln = 4.61$) showing a glass cliff effect, which was more than four standard deviations above the effect size mean. When including this outlier in analyses, the meta-analytic effect size was $OR = 1.74$, $p < .001$. However, because the value had a disproportionate influence on the meta-analytic results (when included, the moderation analyses reported above were no longer statistically significant), we excluded it from the main analyses above.

[16] This finding was the same when including all control moderators in the same regression.

Table 9
Summary of Leadership Suitability Moderation Analyses for Within-Participants Studies

Moderator	*g*	95% CI	*p*	*k*	*Q*	*p*	I^2	*T*
Research group					3.94	.05		0.08
Original	.20	[.05, .35]	.01	6	2.47	.78	0.00	
Other	.02	[−.06, .11]	.60	18	19.43	.30	12.51	
Minority group					0.08	.78		0.09
Women	.07	[−.01, .14]	.10	24	24.21	.39	5.00	
Racial/Ethnic	.03	[−0.22, 0.28]	.82	4	1.13	.77	0.00	
Undergraduate sample					1.38	.24		0.10
Undergraduate	.13	[−.004, .26]	.06	13	12.38	.42	3.07	
Other	.03	[−.07, .13]	.53	11	9.78	.46	0.00	
Manipulation check					1.63	.20		0.08
Successful	.03	[−.06, .12]	.48	16	17.87	.27	16.06	
Unsuccessful or absent	.14	[.004, .27]	.04	8	5.91	.55	0.00	
Reliability					0.59	.44		0.10
High	.08	[−.01, .16]	.07	20	21.45	.31	11.42	
Low or not reported	−.01	[−.23, .20]	.90	4	1.48	.69	0.00	
Type of crisis					4.43	.11		0.07
Feminine crisis	.14	[−.005, .28]	.06	5	2.41	.66	0.00	
Masculine crisis	−.22	[−.52, .08]	.15	1	—	—	—	
Neutral or unclear	.06	[−.02, .15]	.16	18	20.13	.27	15.55	
Leadership history					2.97	.23		0.08
Female leader	−.08	[−.27, .11]	.42	4	0.69	.87	0.00	
Male leader	.07	[−.05, .18]	.24	8	9.92	.19	29.44	
Unknown	.12	[.001, .24]	.05	12	11.73	.38	6.22	
Researcher experience					0.003	.96		0.10
Student	.06	[−.06, .18]	.30	9	8.19	.42	2.32	
PhD student or above	.07	[−.04, .17]	.22	15	14.89	.39	5.98	
Continuous moderators (*B*)								
Gender inequality index	1.49	[0.78, 2.20]	<.001	24	16.94	<.001		0.00
Participant gender	.003	[−.004, .01]	.39	24	0.73	.39		0.10

Note. *g* is the mean Hedge's *g* effect size (bias-corrected) where positive values indicate higher ratings of women or minorities as suitable leaders in crisis situations. For the continuous moderator analysis, the unstandardized *B* coefficient is reported instead, and it describes the relationship between the moderator (gender inequality index of the country under consideration and the percentage of female participants) and the presence of a glass cliff. CI = confidence interval. *k* indicates the number of effect sizes included in each analysis. Q_B tests for differences across subgroups and is reported under *Q* in the first row for each moderator. Q_W tests homogeneity within subgroups and is reported under *Q* in that subgroup's row. Q_R tests whether the moderator accounts for significant variability in the model and is reported under *Q* in the case of continuous moderation analyses. I^2 indicates the percentage of heterogeneity stemming from systematic differences across studies, rather than from sampling error. *T* describes the standard deviation of the true effect sizes across studies. Dashes indicate insufficient degrees of freedom to produce an estimate.

comparing effects with a history of female leadership to an unknown history of leadership.

Both the gender inequality index and the effect of research group remained significant moderators (R^2 = .53), showing that the glass cliff was larger in countries with more gender inequality and for studies conducted by the original authors. Figure 11 visualizes the moderation by gender inequality, suggesting that the glass cliff occurs in countries with a gender inequality index above approximately 0.10. This includes countries such as the United Kingdom, United States, and Thailand, but not Germany and Switzerland. Additionally, the glass cliff was larger when the gender of the previous leader was unknown than when the previous leader was a woman or a man (see Table 12).[17]

Table 10
Summary of Leadership Suitability Multiple Meta-Regression Analysis for Within-Participants Studies

Moderator	*B*	95% CI	β	*p*
Gender inequality index	1.38	[0.63, 2.12]	.66	<.001
Research group	0.08	[−0.08, 0.23]	.18	.32

Note. The multiple meta-regression analysis included all moderators for which $p < .05$ in bivariate analyses. *B* = unstandardized regression coefficient. CI = confidence interval. β = standardized beta coefficient. $k = 24$ and indicates the number of effect sizes included.

Race subanalysis. To explore evidence supporting three of the explanations for the glass cliff, we again tested whether the effect size varied between different racial and ethnic groups. None of the effects were statistically significant (and the number of effect sizes was limited), but we nonetheless present the direction of effects for exploratory purposes. We only found a pattern in the direction of the glass cliff for South Asian leaders, *OR* = 2.25 (k = 1), and Middle Eastern leaders, *OR* = 1.21 (k = 2), but not for Black, *OR* = .88 (k = 2), or East Asian, *OR* = .92 (k = 1), leaders. Once again, this pattern is inconsistent with the suggestion that underrepresented groups are appointed to leadership positions in times of crisis if they are stereotyped with feminine qualities.

[17] When instead using female leader as the reference group, the meta-analytic effect size for female leaders did not differ from those of male leaders, *OR* = 1.43, *p* = .15.

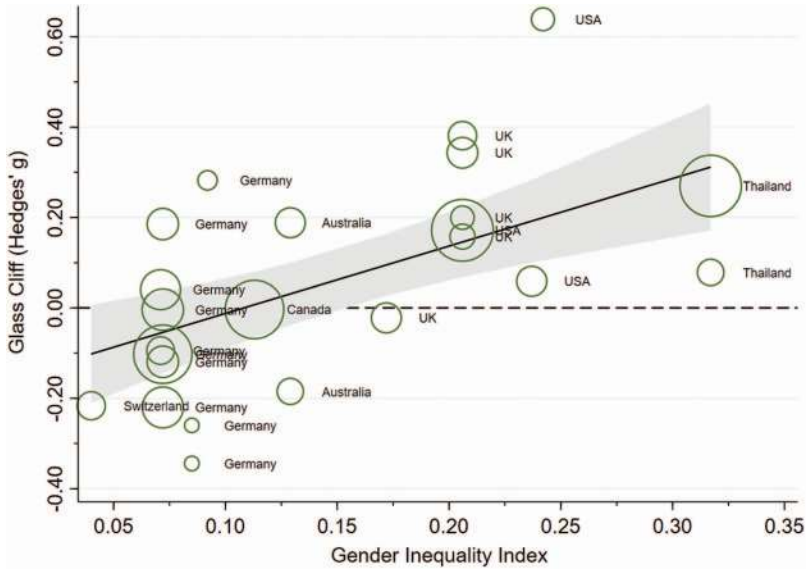

Figure 8. Glass cliff effect for leadership suitability ratings in within-participants studies plotted as a function of the country's gender inequality index. Gender inequality index for a particular country can vary because the index changes from year to year (the index shown is for the year of data collection). Effect sizes above 0 indicate a glass cliff effect (higher ratings of women or minorities as suitable leaders in crisis situations). The size of each diamond reflects the relative weight of the effect in the meta-analysis. The shaded area indicates the 95% confidence band. See the online article for the color version of this figure.

Nevertheless, as this analysis only included a small number of effects, results should be interpreted with caution.

Publication Bias

Our sample included a large proportion of dissertations, theses, or other unpublished research: 50% in the archival analysis, 40% in the between-participants leadership suitability studies, 75% in the within-participants leadership suitability studies, and 63% in the leader selection studies. First, we compared the effect sizes across publication type for each of the analyses. In the archival analysis and both leadership suitability analyses, unpublished work showed a comparable effect size to that of published studies (see Table 13 for statistics). In the experimental studies analysis on leader selection, unpublished work showed a weaker effect than the published studies: The glass cliff effect held for published, but not for unpublished studies, indicating that null results have been published at a lower rate than results showing significant glass cliff effects.

Across our analyses, the large proportion of unpublished research shows that publication bias may be less of a concern than is typical in psychological meta-analyses. We nonetheless conducted analyses assessing publication bias using a selection model because publication bias can rarely be ruled out entirely (also see funnel plots in Figures 3, 5, 7, and 10). Our selection models assumed that both statistically significant and nonsignificant effect sizes may be published, but with different likelihoods. We used a cutoff of $p = .05$ in all analyses, with the exception of the between-participants experimental studies analysis on leadership suitability ratings because the model did not converge at p values of .05 or .01; thus, we used a cutoff of $p = .10$ in this analysis. In the archival analysis, the likelihood ratio test showed no evidence of publication bias, and the bias-adjusted effect size estimate was larger than the original naïve estimate (see Table 13). This finding likely reflects the large proportion of effect sizes included (79%) that were not statistically significant. In all three experimental meta-analyses, the bias-adjusted effect size estimates were slightly smaller than the original naïve estimates, but the likelihood ratio test showed that they were not better fits to the data than the naïve estimates, suggesting no evidence of publication bias.

In sum, given that the majority of the included research was unpublished, the unpublished effect did not substantially differ from the published effect (except for the leader selection experiments), and that the selection models did not indicate publication bias, there is a relatively low likelihood of publication bias in this body of literature.

Discussion

All three of the meta-analyses conducted provided some evidence in line with the glass cliff: The analysis of archival data shows that members of underrepresented groups were more likely to be appointed to leadership positions in times of crisis. The analysis of the experimental data shows that members of underrepresented groups were (a) rated as more suitable for leadership positions (albeit only in between-participants designs) and (b) more likely to be selected for leadership positions in times of crisis. We thus found support for the glass cliff across different methodologies, despite collapsing across some moderators, which likely resulted in a reduction of the size of the effects.

Across the analyses, there was little evidence that publication bias had an impact on the size of the effects, perhaps because of

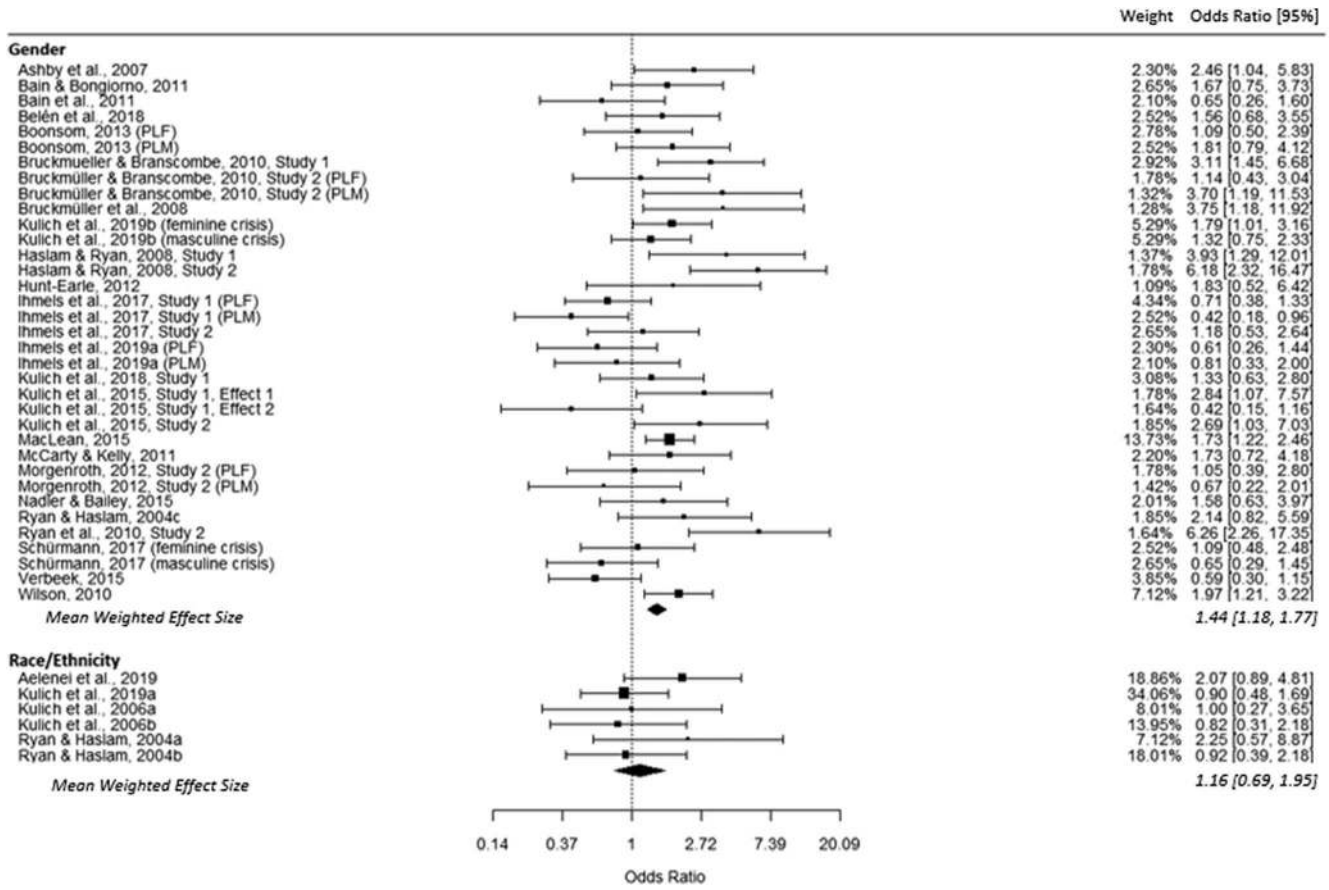

Figure 9. Forest plot of experimental odds-ratio effect sizes (leader selection) with 95% confidence intervals. Effect sizes above 1 indicate a glass cliff effect (increased likelihood of selecting women or minorities for crisis situations). Percentages, as well as the size of each point, reflect the relative weight of the effect. PLF (previous leader female) and PLM (previous leader male) designate independent samples within a study where the previous leader was either female or male (manipulated between-participants), respectively.

our substantial efforts to obtain unpublished studies on the glass cliff. These efforts resulted in a large proportion (40% to 75%, depending on the analysis) of the effect sizes included in our analyses coming from dissertations, theses, or other unpublished research, which is relatively unusual for meta-analyses in psychology.

Taken together, these meta-analyses suggest that there is evidence for the glass cliff phenomenon, but the variability in the data and the impact of moderators points to the need both for a nuanced appreciation of the phenomenon and for future research. We structure this discussion in terms of a discussion of these key moderators, followed by the identification of open questions and implications.

Moderation

The glass cliff has been described as a context-sensitive phenomenon that is likely to arise in particular circumstances at particular times (Ryan & Haslam, 2007; Ryan et al., 2016). For this reason, we would expect the appointment of women to risky and precarious positions to be moderated by several key variables.

Across all analyses of experimental studies, there was no evidence for moderation by sample type (e.g., undergraduate vs. working samples) nor by participant gender, suggesting that the glass cliff is relatively robust across samples, and that it is not simply a case of ingroup bias. Additionally, none of the variables indicating study quality moderated the effects, suggesting that the findings (or lack of findings) are not just based on low quality studies.

There was, however, evidence that the glass cliff was moderated by (a) domain (e.g., management, education, politics), (b) history of leadership, (c) gender inequality in the country where the study was conducted, (d) methodology (within or between participants design), and (e) research group (e.g., original authors or not). We will consider each of these in turn, as well as the findings regarding underrepresented racial and ethnic groups, starting with the findings that can speak to the different explanations for the glass cliff—that women and members of underrepresented racial and ethnic groups are appointed to leadership positions in times of crisis to signal change (*signaling change* explanation), because stereotypically feminine attributes are seen as a good fit with what

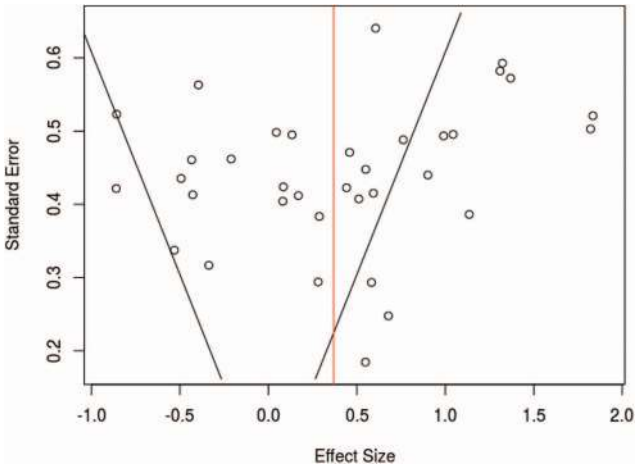

Figure 10. Funnel plot of experimental natural log (ln) odds-ratio effect sizes (leader selection) with black contour lines added at the *p* value cutpoint (.05). Effect sizes located outside of the contour lines represent statistically significant effects, whereas those within the contour lines are not statistically significant. Effect sizes above 0 indicate a glass cliff effect (increased likelihood of selecting women or minorities for crisis situations). The vertical line indicates the meta-analytic mean estimate. See the online article for the color version of this figure.

is needed in times of crisis (*think crisis—think female* explanation), or because they are appointed to risky positions because of *prejudice*.

Evidence for and against the proposed causes of the glass cliff. The results across all three meta-analyses revealed that the glass cliff was apparent in the appointment of members of underrepresented racial and ethnic groups and that these effects were comparable to those for women. The presence of the glass cliff across very different underrepresented groups could be interpreted as support for the idea that organizations appoint underrepresented groups to signal change (Kulich et al., 2015), as any visible underrepresented group can serve this function. Additionally, this finding can be interpreted as support for the idea that both racial and gender prejudice underlies the emergence of the glass cliff.

In contrast, the finding does not support the notion that the glass cliff occurs because stereotypically feminine qualities are seen as useful in times of crisis (e.g., Ryan et al., 2011). Specifically, stereotypes that women are tactful and have a desire to avoid controversy lead people to view them as well suited to deal with organizational issues in times of crisis (Ryan et al., 2011). Given that the same stereotypes do not exist for many underrepresented racial and ethnic groups (see Devine & Elliot, 1995), this account is less plausible as an explanation for a race- or ethnicity-based glass cliff. At the same time, however, Gündemir and colleagues (2019) have demonstrated that the glass cliff only applies to underrepresented racial groups to the extent that the stereotypes of their group are consistent with those of women. Indeed, along the lines of the *think crisis – think female* association, these researchers demonstrated that East Asian Americans were seen as more feminine than other racial groups and in turn were seen as more suitable for leadership positions in times of crisis.

We ran exploratory analyses investigating whether, in line with the findings reported by Gündemir and colleagues (2019), the effect was more pronounced for racial and ethnic groups that are stereotyped as more feminine (e.g., East Asians) than for groups that are stereotyped as more masculine (e.g., Blacks). The meta-analysis results did not map onto masculine and feminine stereotypes and were therefore more in line with the idea that organizations appoint members of underrepresented groups to leadership positions to signal change or due to prejudice. Again, however, the number of studies available for these analyses was small, so the results should be interpreted with caution. It would be helpful to unpack these processes further in future research.

The results from the analyses examining history of leadership were somewhat unclear. Surprisingly, at least in some of the analyses, the glass cliff effect was the strongest when the gender of the previous leader was unknown. Perhaps this result reflects the fact that gender was less salient in these cases and the research question thus less obvious. Yet, given that the number of studies in which it was known was small and previous research has indicated that when no information is given, participants assume that the previous leader was male by default (Kulich et al., 2015), we interpret this as additional, albeit ambiguous, support for the *signaling change* explanation.

In contrast, the results from the archival meta-analysis regarding moderation by domain did not clearly support the *signaling change* or the *prejudice* explanation. We found no evidence for the glass cliff in the management domain—the domain where it was first discussed and observed, and the domain on which most studies have focused—or the sports domain. We found a small effect in the political domain and more substantial effects in the more female-dominated domains (education and nonprofit). Arguably, women would make a better signal for change and face more discrimination in domains in which they are more underrepresented, so this finding raises questions about the signaling change explanation.

The lack of a glass cliff in the management domain is notable because it is the domain in which the effect was originally established. One potential explanation for the null effect could be the increased use of quotas or voluntary targets for women's representation in upper management in some countries (e.g., Norway, Germany, the United Kingdom). This may give companies less leeway to discriminate and disproportionately appoint women to leadership positions during times of crisis. Importantly, even the prominence of discussions of quotas and efforts to avoid them (e.g., by setting voluntary targets) may have similar, albeit weaker, effects on the appointment of women (see also Morgenroth & Ryan, 2018). Quotas, and widespread discussion of quotas, may also explain why the effect for the political domain, where quotas are also common (at least in some countries), was small. Importantly, while we did not consider the presence of quotas as a moderator in our analysis, research suggests that when such quotas are in place for political parties, the glass cliff does not emerge (e.g., Ryan et al., 2010).

Two other moderators were included to examine the *prejudice* explanation: Percentage of male participants (as an indicator of potential ingroup bias) and country-level gender inequality. The percentage of male participants was unrelated to the emergence of the glass cliff, which dovetails with research suggesting that women sometimes perpetuate and justify gender inequality as well (Ridgeway, 2011). We found no evidence that gender inequality moderated the glass cliff effect in the archival meta-analysis or for

Table 11
Summary of Leader Selection Moderation Analyses

Moderator	*OR*	95% CI	*p*	*k*	*Q*	*p*	I^2	*T*
Research group					9.32	.002		0.38
Original	2.42	[1.65, 3.56]	<.001	9	13.01	.11	38.51	
Other	1.22	[0.98, 1.51]	.07	26	23.88	.53	0.00	
Minority group					0.58	.45		0.43
Women	1.44	[1.18, 1.77]	<.001	35	41.08	.19	17.23	
Racial/Ethnic	1.16	[0.69, 1.95]	.57	6	2.25	.81	0.00	
Undergraduate sample					0.64	.42		0.47
Undergraduate	1.53	[1.19, 1.98]	.001	25	32.32	.12	25.74	
Other	1.27	[0.88, 1.85]	.20	10	4.73	.86	0.00	
Manipulation check					0.19	.66		0.47
Successful	1.40	[1.07, 1.82]	.01	22	26.46	.19	20.63	
Unsuccessful or absent	1.54	[1.08, 2.19]	.02	13	10.78	.55	0.00	
Type of crisis					3.14	.21		0.43
Feminine crisis	2.00	[1.27, 3.16]	.003	6	4.20	.52	0.00	
Masculine crisis	0.97	[.45, 2.11]	.95	2	0.79	.37	0.00	
Neutral or unclear	1.37	[1.08, 1.74]	.009	27	32.67	.17	20.42	
Leadership history					7.73	.02		0.39
Female leader	0.79	[0.47, 1.33]	.38	5	2.22	.70	0.00	
Male leader	1.25	[0.84, 1.86]	.27	8	6.05	.53	0.00	
Unknown	1.73	[1.36, 2.21]	<.001	22	28.44	.13	26.16	
Student project					0.81	.37		0.47
Student	1.26	[0.87, 1.82]	.23	11	7.84	.64	0.00	
Not student	1.55	[1.20, 2.00]	<.001	24	29.40	.17	21.77	
Continuous moderators (*B*)								
Gender inequality index	3.62	[1.46, 5.78]	.001	35	10.77	.001		0.34
Participant gender	.004	[−.01, .02]	.64	35	0.22	.64		0.47

Note. *OR* is the mean weighted odds ratio effect size where values above 1 indicate increased likelihood of selecting women or minorities for crisis situations. For the continuous moderator analysis, the unstandardized *B* coefficient is reported instead of the odds ratio, and it describes the relationship between the moderators (gender inequality index and percentage of female participants) and the presence of a glass cliff. CI = confidence interval. *k* indicates the number of effect sizes included in each analysis. Q_B tests for differences across subgroups and is reported under *Q* in the first row for each moderator. Q_W tests homogeneity within subgroups and is reported under *Q* in that subgroup's row. Q_R tests whether the moderator accounts for significant variability in the model and is reported under *Q* in the case of continuous moderation analyses. I^2 indicates the percentage of heterogeneity stemming from systematic differences across studies, rather than from sampling error. *T* describes the standard deviation of the true effect sizes across studies.

between-participants experimental studies. However, in within-participants studies, the glass cliff was larger in countries with higher gender inequality for both leadership suitability ratings and leadership selection. As illustrated in Figures 8 and 11, the glass cliff only emerged for a subset of countries with higher levels of gender inequality, such as the United Kingdom, United States, and Thailand, but not for countries with lower levels of gender inequality such as Germany and Switzerland. This pattern indicates that society-level gender equality plays a part in the emergence of the glass cliff, at least in some circumstances (e.g., when gender is more salient or potential discrimination more obvious, as it would be in within-participants studies). At the same time, the figures also illustrate that some country-level factors beyond gender inequality affect the emergence of the glass cliff. For example, despite low levels of gender inequality, studies conducted in Spain all showed evidence for the glass cliff.

We found no support for the *think crisis—think female* explanation.[18] As mentioned above, the glass cliff effect was not more pronounced for racial and ethnic groups that are stereotyped as more feminine. Moreover, there was no evidence that femininity of crisis moderated the effect. However, the majority of crises were coded as neither clearly feminine nor masculine, so future studies should examine this question more explicitly.

The three explanations are of course not mutually exclusive. It could well be that all three processes feed into the emergence of the glass cliff. However, taken together, our analyses do not find support for the *think crisis—think female* explanation, while finding some (albeit inconsistent and indirect) support for the *signaling change* and the *prejudice* explanations.

Methodology and bias moderators.

Study design. Moderation analyses also revealed that study design affected the emergence and size of the glass cliff effect for leadership suitability ratings. Thus, the effect was present in studies that used a between-participants designs in which participants were asked to rate only a woman or a man, but not in those with within-participant designs where participants evaluated both women and men simultaneously. This difference was despite the fact that materials for the two designs were otherwise extremely similar. It may be that the effect is harder to find in within-participants studies because being exposed to both a woman and man makes the purpose of the study more transparent, and the possibility of bias more obvious, leading participants to put more effort into appearing unbiased. Yet, the fact that an effect emerges

[18] It may seem that our findings regarding domain support the *think crisis—think female* explanation. However, although the risky leadership positions in more female-dominated fields may indeed be more feminine, this would arguably also be true for the nonrisky positions. In other words, it is unclear why crises in particular should require more feminine skills in these domains.

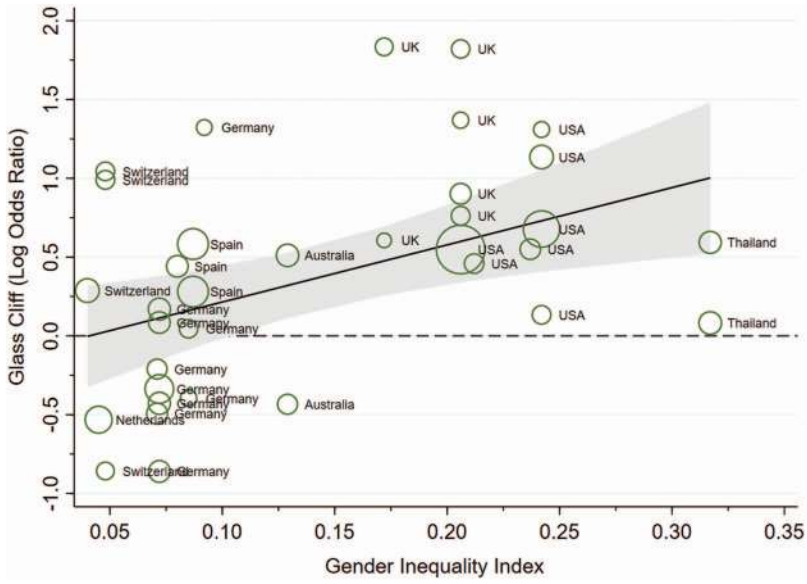

Figure 11. Glass cliff effect (increased likelihood of selecting women or minorities for crisis situations) in natural log (ln) odds ratio for leadership selection in experimental studies plotted as a function of the country's gender inequality index. Gender inequality index for a particular country can vary because the index changes from year to year (the index shown is for the year of data collection). The size of each diamond reflects the relative weight of the effect in the meta-analysis. The shaded area indicates the 95% confidence band. See the online article for the color version of this figure.

in the leadership selection studies may speak against this explanation—by design, these studies all manipulate gender of the candidate within participants, asking them to choose a woman or a man for the position. On the other hand, the forced choice may make it harder not to express bias.

Research group. For the archival and the experimental leadership suitability studies, we did not find a stronger glass cliff in research from one of the two original authors than from other authors. For the leader selection studies, while there was an overall glass cliff effect, it was larger for studies in which at least one of the original authors was involved. Because this was the only analysis for which authorship impacted the effects, the glass cliff literature does not appear to stem exclusively from the findings of the original research group. Nevertheless, it is worth discussing why authorship seems to matter for within-participants studies. One of the reasons could lie in country-level gender inequality of the studies ran by the different research groups. Studies carried out by the original authors primarily used British samples, whereas participants for all other studies come from different countries, with the largest proportion coming from Germany. Notably, our measure of gender inequality, the GII, is consistently lower in Germany than the United Kingdom (see United Nations Development Programme, 2019). In line with this explanation, the effect of research group in within-participants studies no longer emerged when accounting for gender inequality, although the effect did remain for the leader selection studies. It would be beneficial to understand if the glass cliff replicates in a more diverse set of countries and across a wider range of research groups.

Table 12
Summary of Leadership Selection Multiple Meta-Regression Analysis

Moderator	*B*	95% CI	*OR*	*p*
Leadership history				
Male leader	−0.38	[−0.69, −0.08]	0.68	.01
Female leader	−0.74	[−1.18, −0.31]	0.48	<.001
Gender inequality index	3.57	[1.92, 5.21]	35.40	<.001
Research group	0.42	[.10, .73]	1.52	.009

Note. The multiple meta-regression analysis included all moderators for which $p < .05$ in bivariate analyses. *B* = unstandardized regression coefficient. CI = confidence interval. β = standardized beta coefficient. $k = 25$ and indicates the number of effect sizes included. The leadership history variable was dummy coded such that unknown leader was the reference group (0), and other categories were coded as 1.

Open Questions

Although the analyses reported above are an important first step in making sense of the glass cliff literature as a whole, many open questions remain. First, although we found the glass cliff across all three analyses, the effect was small. It is unclear whether this reflects a truly small effect or an effect that is small under some conditions and larger under others. Indeed, there were many moderators investigated across the studies included in the analyses. Unfortunately, we could not analyze these systematically as there were very few studies reporting or manipulating each of these moderators. We chose to collapse across them, even when the authors predicted or demonstrated that the glass cliff emerged under one level of the moderator but not the other. This analytic necessity may have obscured some of the nuances of the glass cliff effect.

Table 13
Summary of Publication Bias Analyses

Meta-analysis	Publication status				Selection model			
	Published ES (*p*)	Unpublished ES (*p*)	Q_B (*p*)	Bias-adjusted ES	Original ES	Likelihood ratio *p*	Lower cutoff *N*	Higher cutoff *N*
Archival	0.02 (.66)	0.11 (.03)	1.45 (.23)	0.13	0.07	.11	6	22
Leader suitability (between)	0.33 (.01)	0.27 (.07)	0.08 (.78)	0.28	0.39	.91	4	6
Leader suitability (within)	0.19 (.04)	0.04 (.41)	2.33 (.13)	0.02	0.07	.17	5	19
Leader selection	2.30 (<.001)	1.16 (.19)	11.23 (<.001)	1.28	1.45	.36	12	23

Note. ES is the effect size. For the first three rows, this corresponds to the mean Hedge's *g* effect size (bias-corrected) where positive values indicate higher ratings of women or minorities as suitable leaders in crisis situations. For the final row, this corresponds to the mean odds ratio effect size where values above 1 indicate increased likelihood of selecting women or minorities for crisis situations. Q_B tests for differences across effect sizes from published and unpublished research studies. The analyses for the selection models assumed that both statistically significant and nonsignificant effect sizes may be published, but with different likelihoods. We used a cutoff of $p = .05$ in all analyses, with the exception of the between-participants leadership suitability studies because the model did not converge at *p* values of .05 or .01; thus, we used a cutoff of $p = .10$ in this analysis. The likelihood ratio *p* value tests the null hypothesis that the bias-adjusted model is not better fit to the data than the naïve model. The lower cutoff *N* indicates the number of effect sizes that were $p < .05$ (or $p < .10$ for the between-participants leadership suitability studies), whereas the higher cutoff *N* indicates the number $p \geq .05$ (or $p \geq .10$).

Because the majority of studies investigated one or more moderators, this approach is likely to result in an effect size that underestimates how large the effect can be under certain conditions. Such potential moderators include severity of crisis (e.g., Ihmels et al., 2019b; Kührt & Hilse, 2012; Schürmann, 2017), ambivalent sexism of participants (Acar & Sümer, 2018; Kührt & Hilse, 2012), and visibility of the organization (Brunner, 2014; Ihmels et al., 2019b; Morgenroth, 2012).

Implications and Conclusions

Our findings have important potential implications for individuals from underrepresented groups who are in, or who are seeking, leadership positions. We found that women, as well as members of underrepresented racial and ethnic groups, were more likely to be appointed to leadership positions in times of crisis across a range of methodologies. Overall, the findings indicate that the glass cliff is an observable phenomenon, but it may only emerge in certain professional domains or other contexts that require further investigation.

The presence of the glass cliff is problematic for several reasons (also see Ryan & Haslam, 2007; Ryan et al., 2016). First, the nature of the glass cliff limits the opportunities available to members of underrepresented groups, even for those who do manage to break through the glass ceiling. The fact that they are more likely to be offered precarious position not only means that their range of opportunities is limited but also that they may be more likely to fail if they take on the leadership position.

Moreover, owing to gender stereotypes, which portray white men as agentic but women and (some) members of underrepresented racial and ethnic groups as lacking agency (e.g., Eagly, Nater, Miller, Kaufmann, & Sczesny, 2010; Galinsky et al., 2013), this failure is more likely to be attributed to situational factors for white men, but to personal failings for members of underrepresented groups. In this way, the precariousness of the glass cliff, and the potentially high risk of failure, runs the risk of reinforcing and perpetuating stereotypes and inequalities in the workplace.

Yet, in line with previous observations (e.g., Ryan & Haslam, 2007; Ryan et al., 2016) these meta-analyses demonstrate that the glass cliff is neither a ubiquitous nor necessarily a consistent phenomenon. Instead, its magnitude is dependent on a range of contextual factors. This nuance in turn has several implications. First, it reinforces the *glass* in the *glass cliff*, underlying the subtlety of the phenomenon and the potential difficulty in recognizing it. Second, the variable strength of the glass cliff is consistent with evidence that the phenomenon is multiply determined, with a range of structural and psychological underpinnings (Ryan & Haslam, 2007; Ryan et al., 2016).

Taken together, we find support for the glass cliff, but the effect is small and not ubiquitous. Instead, it is restricted to some domains and may depend on a range of factors. Interestingly, some of the factors that we expected to impact the magnitude of the glass cliff (e.g., the nature of the crisis measure) did not moderate or did not moderate in the ways we expected (e.g., professional domain), so further research is needed. On the one hand, such findings give us hope. They suggest to us that change is indeed possible and that members of underrepresented groups are not handed leadership positions just to be pushed off the edge of the glass cliff across all leadership contexts. On the other hand, the fact that the glass cliff effect is context-dependent may make it more challenging to tackle the glass cliff where it does exist. A phenomenon that is complex and difficult to identify is unlikely to have an easy solution, but more nuanced research into its mechanisms may help garner support for resources to combat the problem.

References

References marked with an asterisk indicate studies included into one or more meta-analyses.

*Acar, F. P. (2015). Gender differences in promotions to top level management positions: Anexamination of glass cliff in the IT Sector. *Procedia: Social and Behavioral Sciences, 210,* 223–230. http://dx.doi.org/10.1016/j.sbspro.2015.11.362

*Acar, F. P., & Sümer, H. C. (2018). Another test of gender differences in assignments to precarious leadership positions: Examining the moderating role of ambivalent sexism. *Applied Psychology, 67,* 498–522. http://dx.doi.org/10.1111/apps.12142

Adams, S. M., Gupta, A., & Leeth, J. D. (2009). Are female executives over-represented in precarious leadership positions? *British Journal of Management, 20,* 1–12. http://dx.doi.org/10.1111/j.1467-8551.2007.00549.x

*Aelenei, C., Assilamehou-Kunz, Y., Iacoviello, V., & Kulich, C. (2019). *The political glass cliff: When left-wing voters choose minority candidates for hard-to-win seats*. Manuscript in preparation.

*Ahlden, O., & Kollberg, F. (2018). *Female directors' relationship to financial performance* (Unpublished master's thesis). Jönköping University, Sweden.

Andersen, E. G. (2013). *Female CEO's: A study of their appointment performance and market reaction* (Unpublished master's thesis). Aarhus University, Aarhus, Denmark.

*Ashby, J., Ryan, M. K., & Haslam, S. A. (2007). Legal work and the glass cliff: Evidence that women are preferentially selected to lead problematic cases. *William & Mary Journal of Women and the Law, 13,* 775–793.

*Bain, P., & Bongiorno, R. (2011). *Glass cliff study* (Unpublished raw data). University of Queensland, Department of Psychology.

*Bain, P., Bongiorno, R., & Stokman, N. (2011). *Glass cliff study* (Unpublished raw data). University of Queensland, Department of Psychology.

*Bechtoldt, M. N., Bannier, C. E., & Rock, B. (2019). The glass cliff myth?–Evidence from Germany and the U. K. *The Leadership Quarterly, 30,* 273–297. http://dx.doi.org/10.1016/j.leaqua.2018.11.004

Bechtoldt, M. N., Voigt, M., & Bannier, C. E. (2016). *The ceiling trumps the cliff—Gender-related promotion patterns to executive boards in DAX-firms*. Unpublished manuscript.

Belén, E. M. (2018). *¿Las mujeres son percibidas como más idóneas para la gesión de empresas en crisis?: La influencie de los estereotipos de género* [Are women perceived as more suitable for the management of companies in crisis?: The influence of gender stereotypes] (Unpublished student presentation). Jaume I University, Department of Psychology.

Biernat, M., & Fuegen, K. (2001). Shifting standards and the evaluation of competence: Complexity in gender-based judgment and decision making. *Journal of Social Issues, 57,* 707–724. http://dx.doi.org/10.1111/0022-4537.00237

*Boonsom, A. (2013). *Causal effects of organizational performance on choosing of leader gender and perceived suitability of leader as moderated by trait and history of leadership* (Unpublished master's thesis). Chulalongkorn University, Bangkok, Thailand.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Hoboken, NJ: Wiley. http://dx.doi.org/10.1002/9780470743386

*Bowles, B. B. (2013). *The glass cliff: An examination of the female superintendency in South Carolina* (Unpublished doctoral dissertation). University of South Carolina, SC, U. S.

Brady, D., Isaacs, K., Reeves, M., Burroway, R., & Reynolds, M. (2011). Sector, size, stability, and scandal: Explaining the presence of female executives in Fortune 500 firms. *Gender in Management, 26,* 84–105. http://dx.doi.org/10.1108/17542411111109327

*Brown, E. R., Diekman, A. B., & Schneider, M. C. (2011). A change will do us good: Threats diminish typical preferences for male leaders. *Personality and Social Psychology Bulletin, 37,* 930–941. http://dx.doi.org/10.1177/0146167211403322

Browning, E. (2017). *The political glass cliff: Potential causes of female underrepresentation in the U.S. House of Representatives* (Unpublished Honors Program Project). Bourbonnais, IL: Olivet Nazarene University.

Bruckmüller, S. (2007). *Gender stereotypes and the glass cliff: Reduced importance of male attributes results in more frequent selection of female leaders in times of crisis* (Unpublished master's thesis). University of Kansas, Lawrence, KS.

*Bruckmüller, S., Barthel, M., Bzduch, S., Dreisbach, M., & Straubmeier, M. (2008). *Den Ursachen auf der Spur: Die Rolle geschlechtsstereotyper Eigenschaften für den Glass Cliff Effekt* [Tracking down the causes: The role of gender stereotypical attributes for the glass cliff effect] (Unpublished field research practical report). Friedrich-Alexander-Universität, Erlangen-Nürnberg, Germany.

*Bruckmüller, S., & Branscombe, N. R. (2010). The glass cliff: When and why women are selected as leaders in crisis contexts. *British Journal of Social Psychology, 49,* 433–451. http://dx.doi.org/10.1348/014466609X466594

Brunner, A. (2014). *Fähigkeitsunterschiede als Hauptursache des Glass-Cliff-Phänomens?* [Differences in abilities as the main cause of the glass cliff phenomenon?] (Unpublished bachelor's thesis). Technische Universität Dresden, Dresden, Germany.

Burton, L., & Grappendorf, H. (2015, June 5). *Only in crisis? Leadership selection in intercollegiate athletics after an ethical scandal*. Paper presented at the North American Society for Sport Management Conference, Ottawa, Canada.

Carroll, W., Hennessey, S. M., & MacDonald, R. (2013, October). Is there a glass cliff? Examining the phenomenon using board of director appointments in Canada. In Allied Academies. (Eds.) *Proceedings of Academy of Organizational Culture, Communications and Conflict* (pp. 9–13). Arden, NC: Jordan Whitney Enterprises.

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2018, July 31). *Correcting for bias in psychology: A comparison of meta-analytic methods*. Retrieved from osf.io/rf3ys

Catalyst. (1995, March 31). *1995 Catalyst census: Female board directors in the Fortune 500*. Retrieved from https://www.catalyst.org/research/1995-catalyst-census-of-women-board-directors-of-the-fortune-500/

Catalyst. (2019, January 16). *Missing pieces report: The 2018 board diversity census of women and minorities on Fortune 500 boards*. Retrieved from https://www.catalyst.org/wp-content/uploads/2019/01/missing_pieces_report_01152019_final.pdf

Çelebi, E. B., & Saatci, E. Y. (2016). Do crises change the perceived performance of female leaders versus male leaders from the standpoint of junior level employees? *Journal of Management and Strategy, 7,* 45–58. http://dx.doi.org/10.5430/jms.v7n1p45

*Chambers, K. (2011). *The glass cliff: The contribution of social identity and gender stereotypes in predicting leadership preference and trust* (Unpublished doctoral dissertation). Carleton University, Ontario, Canada.

*Cook, A., & Glass, C. (2013). Glass cliffs and organizational saviors: Barriers to minority leadership in work organizations. *Social Problems, 60,* 168–187. http://dx.doi.org/10.1525/sp.2013.60.2.168

Cook, A., & Glass, C. (2014a). Above the glass ceiling: When are women and racial/ethnic minorities promoted to CEO? *Strategic Management Journal, 35,* 1080–1089. http://dx.doi.org/10.1002/smj.2161

*Cook, A., & Glass, C. (2014b). Analyzing promotions of racial/ethnic minority CEOs. *Journal of Managerial Psychology, 29,* 440–454. http://dx.doi.org/10.1108/JMP-02-2012-0066

Cook, A., & Glass, C. (2014c). Women and top leadership positions: Towards an institutional analysis. *Gender, Work and Organization, 21,* 91–103. http://dx.doi.org/10.1111/gwao.12018

Del Prete, S., & Stefani, M. L. (2015). Women as 'gold dust': gender diversity in top boards and the performance of Italian banks. Bank of Italy Temi di Discussione (Working Paper) No, 1014.

Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton trilogy revisited. *Personality and Social Psychology Bulletin, 21,* 1139–1150. http://dx.doi.org/10.1177/01461672952111002

Dobbin, F., & Jung, J. (2011). Board diversity and corporate performance: Filling in the gaps: Corporate board gender diversity and stock performance: The competence gap or institutional investor bias. *North Carolina Law Review, 89,* 809–839.

*Dula, L. A. (2018). *Gender and the governance and performance of nonprofit organizations: Women and the United Way* (Unpublished PhD thesis). Indiana University, Bloomington, IN.

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2010). Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. *American Psychologist*. Advance online publication. http://dx.doi.org/10.1037/amp0000494

Eagly, A. H., & Wood, W. (2012). Social role theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 458–476). Thousand Oaks, CA: Sage Ltd. http://dx.doi.org/10.4135/9781446249222.n49

Elsaid, E., & Ursel, N. D. (2018). Re-examining the glass cliff hypothesis using survival analysis: The case of female CEO tenure. *British Journal of Management, 29,* 156–170. http://dx.doi.org/10.1111/1467-8551.12241

EWOB. (2018, November 29). European Gender Diversity Index (GDI) 2018. Retrieved from https://europeanwomenonboards.eu/wp-content/uploads/2018/11/ewob_facsheet.pdf

*Funk, K. D. (2017). *The causes and consequences of women's representation in local governments* (Unpublished doctoral dissertation). Texas A&M University, College Station, USA.

Gabarrot, F., Bry, C., De Oliveira, P., & Dietz, J. (2014, February). *Gender projection and the glass cliff phenomenon*. Paper presented at the 15th Annual Meeting of the Society for Personality and Social Psychology, Austin, USA.

Galinsky, A. D., Hall, E. V., & Cuddy, A. J. (2013). Gendered races: Implications for interracial marriage, leadership selection, and athletic participation. *Psychological Science, 24,* 498–506. http://dx.doi.org/10.1177/0956797612457783

*Gazijeva, F. (2017). *The role of women selected for glass cliff positions: Are they a signal for change or actual change carriers?* (Unpublished master's thesis). University of Geneva, Geneva, Switzerland.

*Gündemir, S., Carton, A. M., & Homan, A. C. (2019). The impact of organizational performance on the emergence of Asian American leaders. *Journal of Applied Psychology*. Advance online publication. http://dx.doi.org/10.1037/apl0000347

*Haslam, S. A., & Ryan, M. K. (2008). The road to the glass cliff: Differences in the perceived suitability of men and women for leadership positions in succeeding and failing organizations. *The Leadership Quarterly, 19,* 530–546. http://dx.doi.org/10.1016/j.leaqua.2008.07.011

*Haslam, S. A., Ryan, M. K., Kulich, C., Trojanowski, G., & Atkins, C. (2010). Investing with prejudice: The relationship between women's presence on company boards and objective and subjective measures of company performance. *British Journal of Management, 21,* 484–497. http://dx.doi.org/10.1111/j.1467-8551.2009.00670.x

Hedges, L. V. (1984). Advances in statistical methods for meta-analysis. *New Directions for Program Evaluation, 1984,* 25–42. http://dx.doi.org/10.1002/ev.1376

Hedges, L. V., & Vevea, J. L. (1996). Estimating effect size under publication bias: Small sample properties and robustness of a random effects selection model. *Journal of Educational and Behavioral Statistics, 21,* 299–332. http://dx.doi.org/10.3102/10769986021004299

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

*Hennessey, S., MacDonald, K., & Carroll, W. (2014). Is there a glass cliff or a solid ledge for female appointees to the board of directors? *Journal of Organizational Culture, Communications, and Conflict, 18,* 125–139.

*Hirschfeld, M., Ihmels, A., & Wegge, J. (2018). *Think crisis—Think feminine: Effects of gender-specific visual cues on the glass cliff phenomenon*. Unpublished manuscript, Technische Universität Dresden, Germany.

*Hunt-Earle, K. (2012). Falling over a glass cliff: A study of the recruitment of women to leadership roles in troubled enterprises. *Global Business and Organizational Excellence, 31,* 44–53. http://dx.doi.org/10.1002/joe.21441

*Ihmels, A., Shemla, M., & Wegge, J. (2017). *Glass cliff study* (Unpublished raw data). Dresden University of Technology, Department of Psychology.

*Ihmels, A., Shemla, M., & Wegge, J. (2019a). *Gender quotas influence the appointment of women to precarious leadership positions: Archival evidence for the glass cliff in Spain and Germany*. Manuscript submitted for publication.

*Ihmels, A., Shemla, M., & Wegge, J. (2019b). *Signal, fit or both? Integrating the think crisis—Think female and the signaling change approach for glass cliff appointments*. Manuscript in preparation.

Johnson, B. T., & Huedo-Medina, T. B. (2011). Depicting estimates using the intercept in meta-regression models: the moving constant technique. *Research Synthesis Methods, 2,* 204–220.

Judge, E. (2003, November 11). "Women on board: Help or hindrance?," *The Times*, p. 21.

*Kelly, J. T., Brown, E. R., Diekman, A. B., & Schneider, M. C. (2018). The change we believe in: The role of socioeconomic conditions in evaluations of black political candidates. *Electoral Studies, 54,* 254–260. http://dx.doi.org/10.1016/j.electstud.2018.04.008

*Kleineberg, K. (2018). *Spieglein, Spieglein an der Wand: Wer regiert in diesem Land? Der Einfluss der geschlechtsspezifischen äußeren Erscheinung auf das politische Glass Cliff* [Mirror, mirror on the wall: Who rules in this land? The effect of gender stereotypical appearance on the political glass cliff] (Unpublished master's thesis). Technische Universität Dresden, Dresden, Germany.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15,* 155–163. http://dx.doi.org/10.1016/j.jcm.2016.02.012

*Kührt, C., & Hilse, L. (2012). *Einfluss der Krisenstärke auf die Entstehung des Glass Cliff-Effekts im akademischen Kontext* [Effect of the severity of crisis on the emergence of the glass cliff effect in an academic context] (Unpublished term paper). Technische Universität Dresden, Dresden, Germany.

*Kulich, C., Aelenei, C., Assilaméhou-Kunz, Y., & Iacoviello, V. (2019). Experimental_polGC2a. Unpublished data.

*Kulich, C., Gartzia, L., & Komarraju, M. (2019). *Contextualizing the think crisis-think female stereotype: When are female leaders really preferred?* Manuscript in preparation.

*Kulich, C., Iacoviello, V., & Lorenzi-Cioldi, F. (2018). Solving the crisis: When agency is the preferred leadership for implementing change. *The Leadership Quarterly, 29,* 295–308. http://dx.doi.org/10.1016/j.leaqua.2017.05.003

*Kulich, C., Lorenzi-Cioldi, F., Iacoviello, V., Faniko, K., & Ryan, M. K. (2015). Signaling change during a crisis: Refining conditions for the glass cliff. *Journal of Experimental Social Psychology, 61,* 96–103. http://dx.doi.org/10.1016/j.jesp.2015.07.002

*Kulich, C., Ryan, M. K., & Haslam, S. A. (2006a). *U. K. Glass cliff Study 1* (Unpublished raw data). University of Exeter, Department of Psychology.

*Kulich, C., Ryan, M. K., & Haslam, S. A. (2006b). *U. K. Glass cliff Study 2* (Unpublished raw data). University of Exeter, Department of Psychology.

*Kulich, C., Ryan, M. K., & Haslam, S. A. (2012). *U. K. Glass cliff study* (Unpublished raw data). University of Geneva, Department of Psychology.

*Kulich, C., Ryan, M. K., & Haslam, S. A. (2014). The political glass cliff: Understanding how seat selection contributes to the underperformance of ethnic minority candidates. *Political Research Quarterly, 67,* 84–95. http://dx.doi.org/10.1177/1065912913495740

Kurt, D. A. (2011). Glass cliff in relation to hostile and benevolent Sexism (Unpublished master's thesis). Middle East Technical University, Ankara, Turkey.

Lipsey, M. W., & Wilson, D. (2000). *Practical meta-analysis* (applied social research methods). Thousand Oaks, CA: SAGE Publishing.

Lipsey, M. W., & Wilson, D. (2001). *Practical meta-analysis*. Thousand Oaks, CA: SAGE Publishing.

MacDonald, K. (2011). *Examining the "Glass Cliff" phenomenon: A study replication using a Canadian context* (Unpublished master's thesis). University of Prince Edward Island, Charlottetown, Canada.

*MacLean, M. (2015). *Crisis in the pipeline: Appointing women in potentially precarious positions* (Unpublished bachelor's thesis). The University of Queensland, Brisbane, Australia.

*McCarty, M. K., & Kelly, J. R. (2011, January). *The hostile glass cliff: Examining the role of ambivalent sexism in the glass cliff effect*. Paper presented at the 12th Annual Meeting of the Society for Personality and Social Psychology, San Antonio, USA.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22,* 276–282. http://dx.doi.org/10.11613/BM.2012.031

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science, 11,* 730–749. http://dx.doi.org/10.1177/1745691616662243

*Mile, A., & Undheim, A. K. (2018). *Kvinner på glassklippen: En empirisk studie av sammenhengen mellom selskapers prestasjon og nsettelse av kvinnelige daglige ledere* [Women on the Glass Cliff: An empirical study of the relationship between companies' performance and the appointment of female CEOs] (Unpublished master's thesis). Norwegian School of Economics, Bergen, Norway.

*Morgenroth, T. (2012). *Signaling change: What companies hope to gain from appointing women to glass cliff positions* (Unpublished diploma thesis). Technische Universität Dresden, Dresden, Germany.

Morgenroth, T., & Ryan, M. K. (2018). Quotas and affirmative action: Understanding group-based outcomes and attitudes. *Social and Personality Psychology Compass, 12,* e12374. http://dx.doi.org/10.1111/spc3.12374

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105–125. http://dx.doi.org/10.1037/1082-989X.7.1.105

Mulcahy, M., & Linehan, C. (2014). Females and precarious board positions: Further evidence of the glass cliff. *British Journal of Management, 25,* 425–438. http://dx.doi.org/10.1111/1467-8551.12046

*Nadler, J. T., & Bailey, S. (2015). Group discussions and the glass cliff context: An exploratory study of gender and leadership. *North American Journal of Psychology, 17,* 617–632.

O'Brien, D. Z. (2015). Rising to the top: Gender, political performance, and party leadership in parliamentary democracies. *American Journal of Political Science, 59,* 1022–1039. http://dx.doi.org/10.1111/ajps.12173

*Oelbaum, Y. (2016). *Understanding the glass cliff effect: Why are female leaders being pushed toward the edge?* (Unpublished doctoral dissertation). City University of New York, New York.

Peterson, H. (2016). Is managing academics "women's work"? Exploring the glass cliff in higher education management. *Educational Management Administration & Leadership, 44,* 112–127. http://dx.doi.org/10.1177/1741143214563897

*Poppe, L., & Wuttig, F. (2011). *Erweiterung des Glass Cliff Phänomens: Gründe für die Bevorzugung von Minderheiten in riskanten politischen Positionen* [Extension of the glass cliff phenomenon: Reasons for the preference for minorities for risky political positions] (Unpublished term paper). Technische Universität Dresden, Dresden, Germany.

Ridgeway, C. L. (2011). *Framed by gender: How gender inequality persists in the modern world*. New York, NY: Oxford University Press.

Rink, F., Ryan, M. K., & Stoker, J. I. (2012). Influence in times of crisis: How social and financial resources affect men's and women's evaluations of glass-cliff positions. *Psychological Science, 23,* 1306–1313. http://dx.doi.org/10.1177/0956797612453115

*Rink, F., Ryan, M. K., & Stoker, J. I. (2013). Social resources at a time of crisis: How gender stereotypes inform gendered leader evaluations. *European Journal of Social Psychology, 43,* 381–392. http://dx.doi.org/10.1002/ejsp.1954

*Robinson, S. (2019). *The glass cliff in American politics: Electoral risks faced by women candidates for state legislative office in the United States from 2011 to 2016* (Unpublished master's thesis). University of Geneva, Geneva, Switzerland.

*Ryan, M. K., & Haslam, S. A. (2004a). *Unpublished data, Study 1* (Unpublished raw data). University of Exeter, Department of Psychology.

*Ryan, M. K., & Haslam, S. A. (2004b). *Unpublished data, Study 2* (Unpublished raw data). University of Exeter, Department of Psychology.

*Ryan, M. K., & Haslam, S. A. (2004c). *Unpublished data, Study 3* (Unpublished raw data). University of Exeter, Department of Psychology.

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

*Ryan, M. K., & Haslam, S. A. (2007). *Unpublished data, Study 4* (Unpublished raw data). University of Exeter, Department of Psychology.

Ryan, M. K., & Haslam, S. A. (2007). The glass cliff: Exploring the dynamics surrounding the appointment of women to precarious leadership positions. *The Academy of Management Review, 32,* 549–572. http://dx.doi.org/10.5465/amr.2007.24351856

Ryan, M. K., & Haslam, S. A. (2009). Glass cliffs are not so easily scaled: On the precariousness of female CEOs' positions. *British Journal of Management, 20,* 13–16. http://dx.doi.org/10.1111/j.1467-8551.2008.00598.x

Ryan, M. K., Haslam, S. A., Hersby, M. D., & Bongiorno, R. (2011). Think crisis-think female: The glass cliff and contextual variation in the think manager-think male stereotype. *Journal of Applied Psychology, 96,* 470–484. http://dx.doi.org/10.1037/a0022133

Ryan, M. K., Haslam, S. A., & Kulich, C. (2010). Politics and the glass cliff: Evidence that women are preferentially selected to contest hard-to-win Seats. *Psychology of Women Quarterly, 34,* 56–64. http://dx.doi.org/10.1111/j.1471-6402.2009.01541.x

Ryan, M. K., Haslam, S. A., Morgenroth, T., Rink, F., Stoker, J., & Peters, K. (2016). Getting on top of the glass cliff: Reviewing a decade of evidence, explanations, and impact. *The Leadership Quarterly, 27,* 446–455. http://dx.doi.org/10.1016/j.leaqua.2015.10.008

Sabharwal, M. (2015). From glass ceiling to glass cliff: Women in senior executive service. *Journal of Public Administration: Research and Theory, 25,* 399–426. http://dx.doi.org/10.1093/jopart/mut030

Schein, V. E. (1973). The relationship between sex role stereotypes and requisite management characteristics. *Journal of Applied Psychology, 57,* 95–100. http://dx.doi.org/10.1037/h0037128

*Schürmann, V. (2017). *Think crisis—think female? Eine Untersuchung zum Einfluss der Krisenart auf das Glass Cliff Phänomen* [Think crisis - think female? An investigation of the effect the type of crisis has on the glass cliff phenomenon] (Unpublished master's thesis). Technische Universität Dresden, Dresden, Germany.

*Smith, A. E. (2014). Getting to the helm: Women in leadership in federal regulation. *Public Organization Review, 14,* 477–496. http://dx.doi.org/10.1007/s11115-013-0240-0

*Smith, A. E. (2015). On the edge of a glass cliff: Women in leadership in local government. *Public Administration Quarterly, 39,* 484–517.

*Smith, A. E., & Monaghan, K. R. (2013). Some ceilings have more cracks: Representative bureaucracy in federal regulatory agencies. *The American Review of Public Administration, 43,* 50–71. http://dx.doi.org/10.1177/0275074011420997

Stokman, N. (2011). *Humanness and the glass cliff: Are leaders appointed to failing organisations because they are more uniquely human?* (Unpublished bachelor's thesis). University of Queensland, Brisbane, Australia.

*Sun, S. L., Zhu, J., & Ye, K. (2015). Board openness during an economic crisis. *Journal of Business Ethics, 129,* 363–377. http://dx.doi.org/10.1007/s10551-014-2164-1

The World Bank. (2019). *Proportion of seats held by women in national parliaments* (%). Retrieved from https://data.worldbank.org/indicator/SG.GEN.PARL.ZS

Thomas, M., & Bodet, M. A. (2013). Sacrificial lambs, women candidates, and district competitiveness in Canada. *Electoral Studies, 32,* 153–166. http://dx.doi.org/10.1016/j.electstud.2012.12.001

United Nations Development Programme. (2019). *Gender inequality index*. Retrieved from http://hdr.undp.org/en/content/gender-inequality-index-gii

Uyar, E. (2011). *The glass cliff: Differences in perceived suitability and leadership ability of men and women for leadership positions in high and poor performing companies* (Unpublished master's thesis). Middle East Technical University, Ankara, Turkey.

*Verbeek, A. (2015). *Appointing a female CEO as a business strategy: A moderated meditation model* (Unpublished master's thesis). Rotterdam School of Management, Rotterdam, the Netherlands.

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika, 60,* 419–435. http://dx.doi.org/10.1007/BF02294384

*Weber, K. (2016). *The European glass cliff: Does culture matter?* (Unpublished master's thesis). Maastricht University, Maastricht, the Netherlands.

*Wicker, P., Cunningham, G. B., & Fields, D. (2019). Head coach changes in women's college soccer: An investigation of women coaches through the lenses of gender stereotypes and the glass cliff. *Sex Roles, 81,* 797–807. http://dx.doi.org/10.1007/s11199-019-01022-2

*Wilson, D. (2010). *An evaluation of gender leadership preference in the government contracting industry: An examination of the glass cliff phenomenon* (Unpublished doctoral dissertation). Capella University, MN.

Wilson, D. B. (2005). *Meta-analysis macros for SAS, SPSS, and Stata*. Retrieved on January 2, 2014, from http://mason.gmu.edu/~dwilsonb/ma.html

Wilson-Kovacs, M. D., Ryan, M. K., Haslam, S. A., & Rabinovich, A. (2008). Just because you can get a wheelchair in the building doesn't necessarily mean that you can still participate: Barriers to the career advancement of disabled professionals. *Disability & Society, 23,* 705–717. http://dx.doi.org/10.1080/09687590802469198

Appendix

Effect Size Formulas

Effect size calculation for leader selection studies (for 2 × 2 designs comparing subgroups; Morris & DeShon, 2002, p. 114, Table 1):

$$\left(\frac{\overline{X}_{women}-\overline{X}_{men}}{S_{pooled}}\right)_{crisis}-\left(\frac{\overline{X}_{women}-\overline{X}_{men}}{S_{pooled}}\right)_{no\ crisis}$$

Converting point-biserial correlations to Hedges' g (Lipsey & Wilson, 2001, Formula 3.36, p. 62):

$$ES_{sm}=\frac{r_{pb}}{\sqrt{p(1-p)(1-r^2)}}$$

Sampling variance for multiple measures combined as a mean (Borenstein et al., 2009, p. 228, formula 3):

$$\frac{1}{4}(V_1+V_2+2r\sqrt{V_1}\sqrt{V_2})$$

Note: When r was not available for a set of effects, we estimated it using the mean of r in other relevant studies that had multiple measures.

Sampling variance for within-participants studies (single group pretest-posttest in raw score metric; Morris & DeShon, 2002, p. 114):

$$\left[\frac{2(1-\rho)}{n}\right]\left(\frac{n-1}{n-3}\right)\left[1+\frac{n}{2(1-\rho)}\delta_{IG}^2\right]-\frac{\delta_{IG}^2}{[c(n-1)]^2}$$

Sampling variance for between-participants studies (independent groups posttest in raw score metric; Morris & DeShon, 2002, p. 117, Table 2):

$$\left(\frac{1}{\tilde{n}}\right)\left(\frac{N-2}{N-4}\right)(1+\tilde{n}\delta_{IG}^2)-\frac{\delta_{IG}^2}{[c(N-2)]^2}$$

Received October 20, 2018
Revision received March 11, 2020
Accepted March 21, 2020 ■