PART A—Study Materials

Study 1 Questionnaire Items

Expected Gender Discrimination. In Study 1, participants were asked to answer the following six questions on a 7-point Likert scale where 1 = “Strongly Agree” and 7 = “Strongly Disagree”. Items were reverse-coded as needed (values of 1 corresponded to harsher penalties for woman leaders and values of 7 correspond to harsher penalties for leaders who are men) so that they could be combined into an index (α = .75).

1) Women leaders who fail are punished more severely than otherwise similar men leaders who fail.
2) A man who fails in a leadership position is more likely to face reputational losses than an otherwise similar woman who fails in a leadership position.
3) A woman who fails in a leadership position is more likely to be fired than an otherwise similar man who fails in a leadership position.
4) A man who fails in a leadership position is more likely to be demoted than an otherwise similar woman who fails in a leadership position.
5) Women leaders who fail are less likely to be given second chances than otherwise similar men leaders who fail.
6) Men leaders who fail face greater penalties than otherwise similar women leaders who fail.

Studies 2 and 3 Vignette

All participants in Studies 2 and 3 read the following:

Imagine that your company is going to launch a new, unproven initiative that will be unique to your company. No one is sure whether or not it will be successful, but most people at the company believe that it will either be a major flop (costing the company a great deal of money) or a great success (earning the company a great deal of money).

You have the opportunity to be the team leader of this new initiative, if you are willing to leave your team. If the initiative is a success, being the leader of the initiative will have an extremely positive impact on your career.

In the benign failure condition, participants then read the following:

Your current manager is extremely supportive of this opportunity. Not only does your manager encourage you to take it, but also assures you that even if the initiative is a failure, there is always a place on your current team for you.

In the costly failure condition, participants then read the following:

However, it is unclear what will happen to your career if the initiative is not a success. Your current manager has not encouraged you to take it, and has even stated that if the initiative is a failure, there may no longer be a place on the current team for you.

Vignette Manipulation Robustness

As a check of our manipulation's robustness, we analyzed participants' written responses to the above prompts using Linguistic Inquiry and Word Count (LIWC) text analysis software. This empirically validated software (Kahn et al. 2007; Tausczik and Pennebaker 2010) cross-references strings of text with an internal dictionary to provide counts of words and word stems related to a series of subjects – in our case, risk. We use risk because participants should perceive the costly failure condition as more dangerous because they have more to potentially lose. In both Studies 2 and 3, we see that the manipulation produces the intended effect. In Study 2, participants' writings in the costly failure reflected significantly more cognizance of risk (t[255] = -4.030, M = 2.176, SD = 1.239, $p < .001$) than those in the benign failure conditions (M = 1.561, SD = 1.209). In Study 3, participants used significantly more word fragments pertaining to risk when failure in a leadership role was costly (t[482] = -2.928, $p < .01$, M = 2.100, SD = 1.740), compared to when failure was benign (M = 1.690, SD = 1.313).

Study 3 – Additional Measures

Leadership Ambitions. In Study 3, participants rated the probability (on a scale of 0 to 100) that they would take each of the following actions:

1) Suppose the leadership position was directly offered to you. What is the likelihood that you would accept the position?
2) Suppose the leadership position was advertised in your company. What is the likelihood that you would apply for the position?

3) Suppose the leadership position was advertised across the company. What is the probability that you would take actions to increase your likelihood of getting the position? For instance, would you meet with your manager to explain why you would be a good candidate for the position?

These three items were combined in an index of leadership ambitions (= .94)

Anticipated Sanctions for Failure. In Study 3, anticipated sanctions for failure were measured using an index (= .86) based on responses to the prompt, "Suppose your team fails. This costs the organization substantial time and money and the team is disbanded. How likely do you think the following outcomes are?" The outcomes were rated on a probability scale of 0 to 100 as follows:

1) My coworkers will lose respect for me.
2) My manager will lose respect for me.
3) I will be demoted.
4) I will be fired.
5) The whole team working on the initiative will be punished.
6) The whole team working on the initiative will be fired

References

Kahn, Jeffrey H., Renee M. Tobin, Audra E. Massey, and Jennifer Anderson. 2007. "Measuring Emotional Expression with the Linguistic Inquiry and Word Count." *The American Journal of Psychology* 120(2):263-86.

Tausczik, Yla R., and James Pennebaker. 2010. “The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods.” *Journal of Language and Social Psychology* 29(1):24-54.

PART B—USE OF LPMS AND STUDY 2 ANALYSES USING LOGISTIC MODELS

We used linear probability models (LPMs)—OLS regressions that predict binary dependent variables—to estimate the effect of costly failure on participants' leadership ambitions. Since we were primarily interested in the interaction effect between gender and costly failure, an LPM was preferable to a logistic model because, "The magnitude of the interaction effect in nonlinear models does not equal the marginal effect of the interaction term, can be of opposite sign, and its statistical significance is not calculated by standard software," (Ai and Norton 2003:123). Another reason we decided to use an LPM was to make the interpretation of results—and importantly, the comparison of results across studies—more intuitive for the reader. Log-odds and even odds ratios can be highly unintuitive: even when they seem straightforward, odds ratios are prone to misinterpretation (von Hippel 2015). Practitioners are increasingly advocating for the use of LPMs (Angrist and Pischke 2008), as they often predict dichotomous variables just as well as logistic models. Although a common concern is that heteroskedasticity leads linear models to incorrectly estimate p-values for dichotomous variables, Hellevik (2009) made 320 comparisons between the p-values of logistic and linear models and found they were nearly perfectly correlated, $r = .9998$, and found no evidence that one technique systematically provides higher p-values than the other. In addition, the results of the main document are substantively similar to logistic models and predicted probabilities discussed below.

In Table B1, we present the logistic mixed model coefficients from this analysis. Note that to accurately interpret interaction terms in logistic regression, predicted probabilities are required because the interaction term in nonlinear models is often misleading with regards to the direction, statistical significance, and size of an effect (Ai and Norton 2003). For this reason, we present the predicted probabilities (i.e., marginal means) for Models B and C in Figures B1 and B2. As readers can see, costly failure significantly reduces both men and women's leadership ambitions across models, compared to benign failure. Of primary interest is the gender difference under costly failure conditions (i.e., the interaction effect, properly interpreted). Note the effect is statistically significant without controls (p = .033), and marginally significant with controls (p = .053).

TABLE B1: Logistic Mixed Model Coefficients Predicting Leadership Ambitions with Costly Failure X Woman (N = 257).

	Model A	Model B	Model C
Costly Failure[a]	-1.675***	-1.011*	-1.049*
	(.339)	(.487)	(.498)
Woman	-.326	.536	.534
	(.318)	(.588)	(.609)
Costly Failure X Woman		-1.262	-1.215
		(.709)	(.721)
Nonwhite			-.500
			(.677)

Nonwhite X Woman			.070
			(.889)
Associate's Degree			-.448
			(.514)
Bachelor's Degree			.179
			(.443)
Advanced Degree			-.682
			(.511)
Age			.037
			(.083)
Age^2			-.00
			(.00)
Employed			-.360
			(.556)
Intercept	2.270***	1.880***	1.540
	(.331)	(.358)	(1.575)

NOTES: Standard errors appear in parentheses. *$p < .05$. **$p < .01$. ***$p < .001$.

[a] Reference category is benign failure.

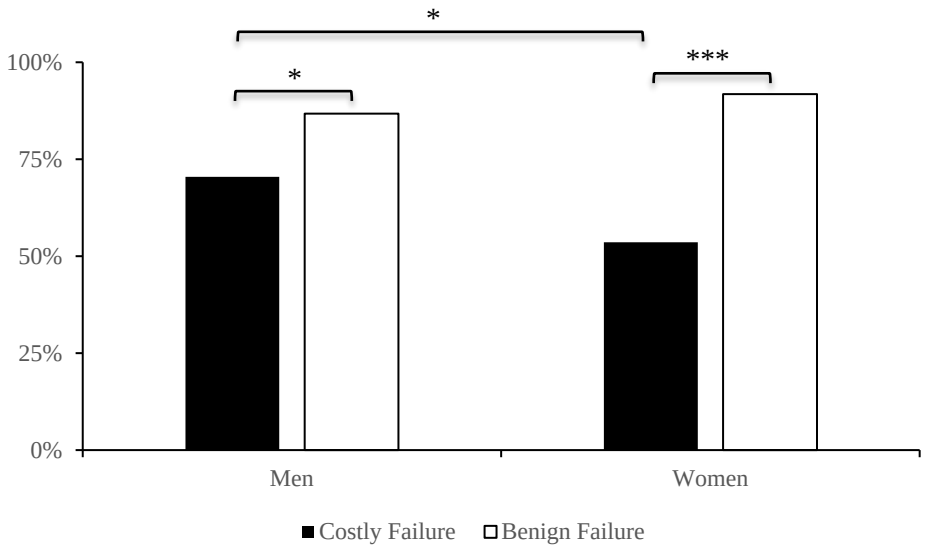

FIGURE B1. Percentage of Men and Women in Costly vs. Benign Failure Conditions Using Predicted Probabilities from Model B.

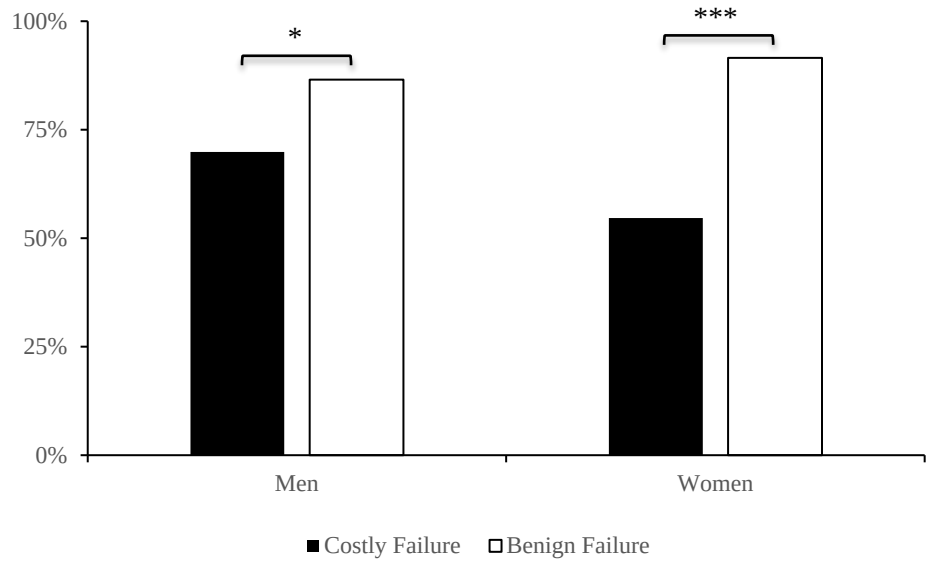

FIGURE B2. Percentage of Men and Women in Costly vs. Benign Failure Conditions Using Predicted Probabilities from Model C.

References

Ai, Chunrong, and Edward C. Norton. 2003. "Interaction terms in logit and probit models." *Economics Letters* 80(1):123–29.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: an Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Hellevik, Ottar. 2009. "Linear Versus Logistic Regression When the Dependent Variable Is a Dichotomy." *Quality and Quantity* 43(1):59-74.

von Hippel, Paul T. 2015. “Linear vs. Logistic Probability Models: Which is Better, and When?” Statistical Horizons. Retrieved June 18, 2018 (https://statisticalhorizons.com/linear-vs-logistic).

PART C—FULL MODELS FOR STUDY 2

TABLE C1: Linear Mixed Model Coefficients Predicting Leadership Ambitions (N = 257).

	Model A	Model B	Model C
Costly Failure[a]	-.288***	-.163*	-.167*
	(.052)	(.079)	(.078)
Woman	-.053	.050	.048
	(.052)	(.072)	(.074)
Costly Failure X Woman		-.219*	-.203*
		(.104)	(.104)
Nonwhite			-.087
			(.110)
Nonwhite X Woman			.001
			(.150)
Associate's Degree			-.076
			(.084)
Bachelor's Degree			.038
			(.067)
Advanced Degree			-.106
			(.080)
Age			.006

			(.013)
Age^2			-.000
			(.000)
Employed			-.047
			(.080)
Intercept	.917***	.868***	.798**
	(.044)	(.050)	(.243)

NOTES: Standard errors appear in parentheses. *$p < .05$. **$p < .01$.

***$p < .001$.

[a] Reference category is benign failure.

PART D—CALCULATION OF INCOME FOR STUDY 3

We asked participants for their annual individual income using response ranges in increments of $10,000 (e.g., “Less than $10,000/year”, “$10,000 - $19,999”, “$20,000 - $29,999”, etc.) until reaching $100,000, at which point we offered the options of “$100,000 - $149,999” and “$150,000+”. Participants could also respond “Don’t know/Refuse to answer.” Following von Hippel, Hunter, and Drown’s (2017) recommendations on modeling binned income categories like the ones we used, we recoded each income category to the value of its midpoint for every category except the top bin. To obtain a robust value for the top bin, we estimated the harmonic mean, obtained by taking the top two bins and fitting Pareto distributions to them, in accordance with equation 1 in von Hippel et al. (2017).

References

von Hippel, Paul T., David J. Hunter, and McKalie Drown. 2017. “Better Estimates from Binned Income Data: Interpolated CDFs and Mean-Matching.” *Sociological Science* 4:641-55. Retrieved June 18, 2018 (https://doi.org/10.15195/v4.a26).

PART E—FULL STUDY 3 MODELS

TABLE E1: Linear Mixed Model Coefficients Predicting Leadership Ambitions (N = 484).

	Model A	Model B	Model C
Costly Failure	-22.930***	-16.539***	-16.622***
	(2.566)	(3.939)	(3.888)
Woman	-3.895	1.551	3.286
	(2.598)	(3.636)	(5.562)
Costly Failure X Woman		-11.043*	-10.982*
		(5.178)	(5.087)
Nonwhite			5.028
			(4.425)
Nonwhite X Woman			-4.481
			(5.688)
Associate's Degree			-2.887
			(3.980)
Bachelor's Degree			-0.675
			(3.408)
Advanced Degree			-1.176
			(4.311)
Age			-1.487*

	(.690)
Age^2	.018*
	(.008)
Employed	4.200
	(4.075)
Log of Income	4.195*
	(1.643)
Married	-.689
	(5.330)
Cohabiting	2.405
	(5.669)
Other Relationship Status	1.453
	(8.360)
Parent	3.132
	(4.751)
Married X Woman	3.736
	(8.037)
Cohabiting X Woman	4.063
	(6.719)
Other Relationship Status X Woman	-10.212
	(10.187)

Parent X Woman			-1.909
			(6.131)
Intercept	78.388***	75.318***	52.620**
	(2.329)	(2.730)	(18.629)

NOTES: Standard errors appear in parentheses. *$p < .05$. **$p < .01$. ***$p < .001$.

[a] Reference category is benign failure.

PART F—OTHER POTENTIAL MEDIATORS IN STUDY 3

Of course, there are other reasons that the costs of failure could amplify gender differences in leadership ambitions. Given existing research on gender and leadership, we test three additional mechanisms that could plausibly create gender differences in leadership ambitions: 1) gender differences in the anticipated odds of failure, 2) gender differences in anticipated resistance from subordinate group members, and 3) gendered social pressures on men to be leaders. While we empirically test these effects for interested readers, we do not position them as our central explanation for why costly failure will disproportionately depress the leadership ambitions of women because these mechanisms are better suited to explain overall gender differences in leadership ambitions, independent of failure's costs.

Anticipated Odds of Failure. Research finds that controlling for task ability, women have lower average levels of self-confidence than men (Niederle and Vesterlund 2007). Indeed, research finds that women rate their ability on male-typed tasks lower than men, even when all participants are given the same performance feedback (Correll 2004). Because women have less confidence in their ability at male-typed tasks, and leadership is a male-typed task, women may believe that they have higher odds of failing in a leadership position, and costly failure could make this belief more salient.

Anticipated Resistance from Subordinates. Women may also be particularly attuned to the costs of failure when they occupy leadership roles because they are concerned that subordinate group members will undermine them, given that women leaders face more resistance from group members than men leaders (Vial, Napier, and

Brescoll 2016). Indeed, a sizeable portion of workers explicitly say they prefer a male over a female boss: 28% of men and 37% of women (Elsesser and Lever 2011).

Pressure from Others to Lead. Lastly, men may feel pressure to accept leadership roles—even when they are subject to high costs of failure—in order to meet norms of masculinity. Hegemonic masculinity states that "ideal" men are leaders (Connell and Messerschmidt 2005), and that successful men should be assertive, business-minded, and risk-taking (Prentice and Carranza 2002). Additionally, men should not be passive, and forgoing a leadership role could be viewed in this manner (Prentice and Carranza 2002).

Methods

Anticipated Odds of Failure. To measure participants' perceptions of their likelihood of failure, we asked them, "Imagine that you became the team leader. What is the likelihood that the initiative would be a great failure?" This was measured using a slider bar scale where 0 = "0 percent chance" and 100 = "100 percent chance".

Anticipated Resistance from Subordinates. To measure participants' expectation that subordinates would undermine them, we asked participants how much they agreed or disagreed with the following questions (where 1 represents "strongly disagree" and 7 indicates "strongly agree"): "If I became the leader, my group members would respect me," (reverse coded) and "If I became the leader, my group members would try to undermine me and my authority." The alpha for these two items was relatively low (α = .50), so we analyzed these two items as individual variables.

Pressure from Others to Lead. To measure participants' expected pressure from others to lead, we asked participants how they thought both 1) their manager and 2) their

co-workers would view the following: a) accepting the leadership position, b) turning down the leadership position, c) applying for the leadership position, d) not applying for the leadership position, e) advocating for the leadership position, and f) not advocating for the leadership position. We measured these items on a 7-point Likert scale where 1 = "extremely negatively" and 7 = "extremely positively." After reverse coding items that tapped into actions against obtaining the leadership position, we combined all 12 items into a single summary score ($\alpha = .88$).

In Table F1, we regress these potential mechanisms on the set of independent variables used in Model C of Table 4 in the main document. If the interaction of gender and costly failure does not significantly predict a proposed mediator, that rules out the mechanism. We find no evidence for these other mechanisms, as neither gender nor the interaction effect (costly failure X woman) is significant in any of these models (Table F1).

TABLE F1: Linear Mixed Model Coefficients Predicting All Discussed Mechanisms (N = 484).

		Anticipated resistance from group members		Pressure from others to lead
	Anticipated odds of failure	*Undermining*	*Disrespect*	(Index)
Costly Failure[a]	3.494	-0.038	0.032	-0.740***
	(2.784)	(0.188)	(0.154)	(0.115)
Woman	2.229	-0.444	-0.367	0.042

	(3.982)	(0.269)	(0.221)	(0.165)
Costly Failure X Woman	-1.330	0.198	0.254	-0.191
	(3.642)	(0.246)	(0.202)	(0.151)
Intercept	55.457***	2.870**	4.300***	5.049***
	(13.338)	(0.901)	(0.740)	(0.552)

NOTE: Standard errors appear in parentheses. *$p < .05$. **$p < .01$. ***$p < .001$.

[a] Reference category is benign failure.

Additional controls are nonwhite, nonwhiteXwoman, associate's degree, bachelor's degree, advanced degree, age, age^2, employed, log of income, married, cohabiting, other relationship status, parent, marriedXwoman, cohabitingXwoman, other relationship statusXwoman, and parentXwoman.

References

Connell, R. W., and James W. Messerschmidt. 2005. "Hegemonic Masculinity: Rethinking the Concept." *Gender & Society* 19(6):829–59.

Correll, Shelley J. 2004. "Constraints into Preferences: Gender, Status, and Emerging Career Aspirations." *American Sociological Review* 69(1):93–113.

Elsesser, Kim M., and Janet Lever. 2011. "Does gender bias against female leaders persist?
Quantitative and qualitative data from a large-scale survey." *Human Relations* 64(12):1555–78.

Niederle, Muriel and Lise Vesterlund. 2007. “Do Women Shy Away from Competition? Do Men Compete Too Much?” *Quarterly Journal of Economics* 122(3): 1067-1101.

Prentice, Deborah A., and Erica Carranza. 2002. “What Women and Men Should Be, Shouldn’t Be, Are Allowed to Be, and Don’t Have to Be: The Contents of Prescriptive Gender Stereotypes.” *Psychology of Women Quarterly* 26(4):269-81.

Vial, Andrea C., Jaime L. Napier, and Victoria L. Brescoll. 2016. “A bed of thorns: Female leaders and the self-reinforcing cycle of illegitimacy.” *The Leadership Quarterly* 27(3):400-14.