Fostering Perceptions of Authenticity via Sensitive Self-Disclosure

Li Jiang
Maryam Kouchaki
Francesca Gino
Reihane Boghrati
Leslie John

Fostering Perceptions of Authenticity via Sensitive Self-Disclosure

Li Jiang
George Washington University

Maryam Kouchaki
Northwestern University

Francesca Gino
Harvard Business School

Reihane Boghrati
University of Pennsylvania

Leslie John
Harvard Business School

Working Paper 20-070

Fostering Perceptions of Authenticity via Sensitive Self-Disclosure

Abstract

Leaders' perceived authenticity – the sense that a leader is acting in accordance with her "true self" – is associated with positive outcomes for both employees and organizations alike. How might a leader foster this impression? Using field and experimental data, we test whether sensitive self-disclosures – for example, revealing an aversion to public speaking – make leaders come across as authentic and lead to positive outcomes, such as an enhanced desire to work with that leader (Studies 1, 2A, 2B, and 3). Stemming from our conceptual account, we show that these benefits emerge when the self-disclosure is mild to moderately sensitive in nature (Study 4) and made voluntarily (as opposed to by requirement) (Study 6) by a relatively high-status person (Study 5). Would-be disclosers do not intuit these positive consequences of self-disclosure (Study 7), suggesting that leaders may under-employ an effective tool for making a positive impression.

Keywords: authenticity, foibles, weaknesses, self-disclosure, leaders, impression

Fostering Perceptions of Authenticity via Sensitive Self-Disclosure

In the 2016 U.S. presidential election, questions of authenticity loomed large (Sergent, 2016; Szalai, 2015; Talbot-Zorn & Marz, 2016; Zimmer, 2016; Zogby, 2016). Donald Trump supporters complained that Hillary Clinton seemed overly rehearsed, even robotic; they found their candidate to be refreshingly candid by comparison. Beyond the political realm, research in organizational behavior indicates that employees prefer leaders they perceive to be authentic (e.g., Clapp-Smith, Vogelgesang, & Avey, 2009; D. S. Wang & Hsieh, 2013; H. Wang, Sui, Luthans, Wang, & Wu, 2014), with Generation Z being particularly likely to prioritize authenticity over other factors when choosing whom to work with (Cronin, 2019; Laudert, 2018).

Consistent with these preferences, perceived authenticity – the perception that leaders are being genuine, acting in accordance with their true selves (Cha et al., 2019; George, Sims, McLean, & Mayer, 2007; Lehman, O'Connor, Kovács, & Newman, 2019) – is associated with positive outcomes for both employees and organizations. When followers perceive leaders to be authentic, they experience greater well-being (Rahimnia & Sharifirad, 2015; H. Wang et al., 2014), are more trusting of the organization (Avolio, Gardner, Walumbwa, Luthans, & May, 2004; Norman, Avolio, & Luthans, 2010), perform better (Hannah, Avolio, & Walumbwa, 2011; Leroy, Palanski, & Simons, 2012; Lyubovnikova, Legood, Turner, & Mamakouka, 2017; Rego, Júnior, & e Cunha, 2015; Rego, Vitória, Magalhães, Ribeiro, & e Cunha, 2013), work harder (Hirst, Walumbwa, Aryee, Butarbutar, & Chen, 2016), and make more ethical decisions (Cianci, Hannah, Roberts, & Tsakumis, 2014; Zhu, Avolio, Riggio, & Sosik, 2011).

Despite these benefits, research also suggests that leaders struggle to come across as authentic (Hahl & Zuckerman, 2014; Hahl, Zuckerman, & Kim, 2017). Leaders are sometimes seen as manipulating their public images to seek power and status – regardless of whether they

are actually engaging in such manipulation – which poses a barrier to being perceived as authentic (Fine, 2003; Hahl & Zuckerman, 2014; Zukin, 2008). Thus, the question arises: What can leaders do to boost followers' perceptions of their authenticity?

Perceived Authenticity

While the construct of authenticity is multi-faceted, scholars have cogently defined its various meanings (Baugh, 1988; Carroll & Wheaton, 2009). Following this prior work, we define authenticity as "consistency between an entity's internal values and its external expressions" (Lehman et al., 2019). In other words, a person is authentic to the extent that her actions – the part of her that is visible to others – are truly reflective of her internal values.

Given this conceptual clarity, at first blush, the answer to the opening question – how can leaders come across as authentic? – might seem to be deceptively simple: just be authentic! But given that the self has many facets, which often conflict with each other (Linville, 1985; Ramarajan, 2014), a new question arises: to be authentic, to which self should one be true? Moreover, given that the "internal self" is unobservable, how can a follower assess whether a leader's external expressions are truly consistent with her internal values?

Perhaps as a result of this fuzziness, scholars have shied away from assessing whether leaders are behaving authentically versus inauthentically in a given situation. In this spirit, we skirt the issue of whether leaders are truly acting authentically, focusing instead on what drives followers' *perceptions* of leaders' authenticity and, more specifically, how leaders may be able to foster such perceptions.

The Role of Sensitive Self-Disclosure in Boosting Perceived Authenticity

By "sensitive self-disclosure," consistent with prior work, we mean information that makes a person vulnerable to being judged negatively by others (Derlega, Metts, Petronio, &

Margulis, 1993; Kelly & McKillop, 1996; Laurenceau, Barrett, & Pietromonaco, 1998; Moon, 2000). Such information tends to be unfavorable, or socially undesirable, self-relevant information, though this definition is naturally context-dependent. For example, a teenager sharing with his parents that he is sexually active would likely be sensitive information, but he probably would be less reticent about revealing this to his friends (in fact, in the context of his peer group, revealing the opposite – virginity – might be sensitive).

People often shy away from disclosing sensitive information (Bruk, Scholl, & Bless, 2018; Gromet & Pronin, 2009; John, Barasz, & Norton, 2016)—sometimes for good reason, given the risks. Disclosure of unsavory or undesirable facts can make a bad impression, given that people tend to overweight negative information relative to positive information (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Herr, Kardes, & Kim, 1991). Indeed, those who give others negative feedback are disliked (Bannister, 1986; Blakely, 1993; Green, Gino, & Staats, 2017; Kingsley Westerman, Reno, & Heuett, 2018), even when they are mere messengers (John, Blunden, & Liu, 2019). Of particular relevance to our paper, revealing weaknesses (e.g., "I am on academic probation", Gibson, Harari, & Marr, 2018) can diminish others' perceptions of the discloser's status.

Perhaps as a result of these negative outcomes, and motivated by self-presentation concerns, people tend to disclose favorable and withhold unfavorable personal information (De Angelis, Bonezzi, Peluso, Rucker, & Costabile, 2012). This is especially true in the workplace, where self-presentation concerns loom large. In job interviews, for example, people engage in extensive image creation, to the point of making up fictional stories to showcase their strengths (Levashina & Campion, 2007).

But we posit that this reticence to reveal weaknesses comes with a cost: it contributes to

perceptions of inauthenticity. If we only reveal our strengths, we are only showing a very narrow "sample" of our true selves. How can we possibly come across as authentic if we only reveal this one side – an obviously biased sample? Thus, we posit that sensitive self-disclosure makes a person come across as relatively authentic, which can in turn produce positive outcomes.

We test this idea as it pertains to leaders in organizations. We focus on leaders because they can accrue organizational benefits from appearing authentic, but at the same time often struggle to come across as authentic. Indeed, leaders tend to be viewed as particularly concerned with impression management (Hahl & Zuckerman, 2014). And, as noted above, sensitive self-disclosure is not without risk. Thus, we ask: How can leaders realize the benefits of disclosure while avoiding its pitfalls?

We posit that leaders can increase their perceived authenticity via self-disclosure, and particularly through self-disclosure of sensitive information. Given that the definition of "sensitive self-disclosure" is context-dependent, how do we define "sensitive self-disclosure" among leaders in the workplace? We argue that when leaders self-disclose their weaknesses, they are engaging in sensitive self-disclosure. This is because disclosing weaknesses, such as an aversion to public speaking, plausibly makes a leader vulnerable to being judged negatively by followers by potentially threatening followers' perceptions of that leader's competence. And to lead effectively, it is important to be perceived as competent (Chng, Kim, Gilbreath, & Andersson, 2018).

Next, we delineate the logic underpinning our predictions for (1) when sensitive self-disclosures will boost perceptions of authenticity, and (2) when this authenticity boost will translate into positive overall impressions and, thus, conditions under which it behooves leaders to reveal their weaknesses.

How Sensitive Is Too Sensitive? Authenticity versus Incompetence

We have posited that when leaders reveal their weaknesses, they come across as more authentic, leading to positive outcomes. However, these potential benefits of revealing weaknesses may be accompanied by potential detriments in perceived competence, which can undermine leaders' ability to lead. When might leaders reap the benefit of revealing their weaknesses (coming across as authentic) and when might the downside (coming across as incompetent) overshadow this benefit?

We advance a "sensitivity sweet spot" hypothesis, such that mild-to-moderately-sensitive disclosures – i.e., a leader's disclosure of her legitimate but not damning weaknesses – produce positive outcomes for leaders. Specifically, we posit that the authenticity gains arising from such disclosures (e.g., revealing one's aversion to public speaking) will more than offset the potential detriment in perceived competence, in turn fostering overall positive impressions from followers. By contrast, we predict that although extremely sensitive disclosures, here in the form of damning weaknesses (such as severe public speaking anxiety), will also increase perceived authenticity, they will erode perceived competence to a greater degree, hurting overall impressions.

Status

We posit that the capacity for self-disclosure of moderately sensitive weaknesses to have positive outcomes to be restricted to high-status individuals, whom we operationalize as leaders in organizations. Specifically, we posit that for high-status individuals, disclosing weaknesses generates pronounced authenticity gains while buffering competence decrements. The rationale for this hypothesis is as follows: First, when it comes to perceived authenticity, given the natural inclination to reveal favorable, as opposed to unfavorable, self-relevant information, revealing

one's weaknesses – a form of unfavorable information – has the capacity to increase the

perceived authenticity of both high- and low-status individuals. However, given that leaders are

particularly concerned with impression management (Chng, Rodgers, Shih, & Song, 2015), and

thus particularly likely to avoid revealing negative self-relevant information, we posit that when

they do reveal such information, they are particularly likely to come across as authentic.

At the same time, we posit that high status buffers against the decrements in perceived

competence that accompany such disclosures because high-status individuals, such as

organizational leaders, are often given the "benefit of the doubt" in terms of competence (Correll

et al., 2017; Lynn, Podolny, & Tao, 2009). Put differently, such individuals are viewed by

default to be relatively competent, and this default perception can offset the competence-eroding

side effect of revealing one's weaknesses (but only up to a point, for, as we note above,

disclosure of damning weaknesses is unlikely to forge positive impressions).

Voluntariness

Finally, we suggest that for leaders' sensitive self-disclosures to boost perceived

authenticity, they must be made *voluntarily*. This prediction stems from the fact that in making

dispositional inferences about a person, observers take intentions into account. For example,

actors are judged to be more moral when they inadvertently, as opposed to intentionally, cause

something bad to happen (Greene et al., 2009). And, of particular relevance to the domain of

self-disclosure, prior work has shown that the negative signal that can arise from explicitly

withholding information (e.g., refusing to answer a direct question) is restricted to situations in

which a person volitionally withholds, as when, for example, they refuse to answer a question (as

opposed to not answering said question simply because they did not notice it) (John et al., 2016).

Analogously, we propose that for leaders to reap the authenticity benefits of revealing sensitive

information, followers must perceive those leaders to be revealing on their own accord. Thus, it is not enough for followers to have awareness of their leaders' foibles: The act of voluntary self-disclosure is crucial to boosting perceptions of authenticity.

Overview of Studies

We begin by presenting field evidence of an association between the sensitivity of leaders' disclosures and employees' positive responses to those revelations (Study 1). Next, in a series of experiments, we show that when leaders disclose their foibles, they come across as authentic, culminating in desirable outcomes, such as followers' positive overall impressions of them and desire to work with and trust them (Studies 2A, 2B, & 3). Stemming from our conceptual account, these benefits emerge when the self-disclosure is mild to moderately sensitive in nature (Study 4), made by a person occupying a high-status role within the organization (Study 5), and made voluntarily (as opposed to by requirement) (Study 6). Finally, we show that would-be disclosers do not intuit these benefits (Study 7), suggesting that leaders may under-employ a tool for creating a positive impression: revealing their foibles.

In testing these ideas, we add to recent work elucidating two cases in which people make better impressions by self-disclosing unglamorous facts, relative to some surprising comparators. First, relative to withholding, disclosing unsavory facts causes a person to be deemed more trustworthy and, in turn, better liked (John et al., 2016). Second, relative to only disclosing one's successes, revealing one's failures can buffer against a negative effect of sharing one's successes: malicious envy (Brooks et al., 2019). Taken together, this research suggests that people's apparent reticence to reveal is sometimes misplaced, and, of particular relevance to our inquiry, may contribute to being perceived as inauthentic.

We pre-specified our sample sizes and did not analyze data until the targets were met. For Study 1, we analyzed all of the data that our field partner gave us. For Studies 2A and 4, we targeted a sample size of 100 participants per cell. For Studies 2B, 5, and 6, we determined our target of 51 participants per cell by a power analysis using the effect size observed in Study 2A (f = .50, α = .05, power = 80%). For our lab studies – Studies 3 and 7 – we collected as much data as we could within a preset number of sessions of data collection. We disclose all manipulations and measures. Data and stimuli are posted at https://osf.io/w24zu/.

Study 1: Field Evidence

Study 1 explores our basic proposition: Leaders can make a good impression on employees via sensitive self-disclosures. We obtained a large dataset from a social-networking platform that allows professionals to connect with other relevant professionals both within their company and across their industry. Unlike many other social-media apps, such as LinkedIn, on which users typically present idealized portraits of themselves, on this platform, people "get real". The platform is designed as a "safe space" that allows users to feel comfortable interacting with candor – to ask difficult questions, give and get advice, vent, or just crack jokes. As a result, this platform presents an ideal context in which to explore our basic proposition, for in addition to housing common, non-revelatory, and even self-promoting posts as on LinkedIn, self-disclosive posts are also prevalent.

We obtained a large anonymized dataset of posts and comments from senior leaders on this platform and the reactions (i.e., "likes") that these posts garnered. Using machine learning, we trained a classifier on a set of human-coded data and predicted labels for the remaining set of the data. In other words, we first had human coders code a small subset of posts and comments for the presence versus absence of self-disclosive content. Next, we trained a machine-learning

algorithm to categorize the remaining sample as either self-disclosive or non-self-disclosive. We then tested whether reactions differed as a function of whether the content was self-disclosive. We hypothesized that self-disclosive content would garner more positive reactions relative to less disclosive content.

Method

      Data. Our dataset consists of posts, comments, and reactions from senior leaders on a professional social-networking platform. As illustrated in Figure 1, which is a screen shot of the platform, a post is the initiation of a new topic or thought. As on Facebook, fellow users can then respond to that post in one of two ways: by commenting and/or by reacting. A comment is a written response to the post; a reaction involves pressing a button to choose one of five possible reactions, all of which are positive: "like" (the default), "helpful," "funny," "uplifting," or "smart." This social network is exclusive, in that it only grants access to those employed at certain firms within consulting and advertising. We requested the posts and comments from all senior leaders (operationalized as Director and above) within consulting firms. We focused on consulting firms because consulting has a stricter hierarchy than advertising does.

      Our dataset consisted of 1,484 posts, 33,589 comments, and 159,221 reactions. The vast majority (93%) of reactions were "likes"; the remaining 7% were distributed as follows: 6.1% "funny," 0.98% "smart," 0.73% "helpful," and 0.73% "uplifting." The number of reactions garnered by any given post or comment ranged from zero to 1,121; however, most (78%) garnered between zero and five reactions. Posts generally received more reactions than comments: the modal number of reactions was one for posts and zero for comments; the average number of reactions was nine for posts and four for comments.

      Procedure. First, one of the authors worked iteratively with three research assistants to

develop a simple binary coding scheme to code all (1,484) posts and a random selection of 2,000

comments. Hereafter we refer to these posts and comments as "observations." This quantity of

observations was deemed large enough to train a machine-learning algorithm, yet reasonable for

human coders to code manually. Observations containing sensitive self-disclosure were coded as

1; those not displaying such disclosure were coded as 0. Per prior work, we defined sensitive

self-disclosure as information that made the discloser vulnerable to being judged negatively by

others (c.f. Derlega et al., 1993; Kelly & McKillop, 1996; Laurenceau et al., 1998; Moon, 2000).

The human coding process went as follows: The team of four coders independently coded

approximately ten randomly selected observations, resolved disagreements by discussion, and

used that discussion to build a refined understanding of what, within this context, qualifies as a

sensitive self-disclosure (see Table 1 for examples). The team repeated this process twice for a

total of approximately 30 randomly selected observations. Then two of the research assistants

independently coded approximately 50 additional randomly selected observations. Their

agreement rate was 79.4%. Disagreements were resolved via discussion. One of the research

assistants then coded the remaining ~3,400 observations.

Next, we used a randomly selected sample of 80% of the 3,484 human-coded

observations to train a classifier capable of predicting labels for the remaining 31,589

observations that had not been human-coded. To do so, we used BERT (Bidirectional Encoder

Representations for Transformers, Devlin, Chang, Lee, & Toutanova, 2018), the state-of-the-art

deep-learning model in natural language processing. BERT learns contextual relations between

words in text data. When used as a classifier, BERT adds a neural layer on top of the base model

and predicts a label for a given input text.

We tested the predictive validity of the classifier by applying it to a holdout sample of

20% of the 3,484 human-coded observations. The classifier achieved 96% accuracy in this

holdout sample – i.e., for 96% of observations, the classifier's categorization agreed with that of

the human-coder classification.

Results. Posts and comments categorized as self-disclosive garnered significantly more

(positive) reactions relative to non-self-disclosive ones ($M_{disclosive}$ = 6.98, $M_{non-disclosive}$ = 4.44,

$t(1529.1)$ = -5.89, $p < .001$); specifically, they garnered more likes ($M_{disclosive}$ = 6.41, $M_{non-disclosive}$

= 4.06, $t(1517.1)$ = -5.94, $p < .001$). Other reactions, though used relatively infrequently, were

also more prevalent for self-disclosive observations; specifically, such observations were more

likely to be deemed helpful ($M_{disclosive}$ = 0.09, $M_{non-disclosive}$ = 0.04, $t(1477.9)$ = -3.41, $p < .001$) and

uplifting ($M_{disclosive}$ = 0.10, $M_{non-disclosive}$ = 0.03, $t(1457.8)$ = -3.61, $p < .001$). The prevalence of

funny reactions was not significantly different as a function of observation type ($M_{disclosive}$ = 0.32,

$M_{non-disclosive}$ = 0.27, $t(1727.3)$ = -0.94, $p = 0.35$). These results hold when analyzing the two types

of observations – posts versus comments – separately (see Supplemental Materials).

In sum, consistent with our basic hypothesis – that leaders' sensitive self-disclosures can

garner positive reactions – these data reveal a positive association between sensitive self-

disclosure and positive reactions. Next, we turn to experiments to test for a causal relationship.

Studies 2A & 2B: Causality

Studies 2A and 2B experimentally test the idea that when leaders disclose their foibles,

they come across as authentic, culminating in positive outcomes, such as making a positive

impression on followers (Study 2A) and increasing followers' desire to work with them (Study

2B). Both studies employed the same paradigm: Participants imagined that during a job

interview, a CEO disclosed information to them. Between subjects, we varied the nature of the

CEO's disclosures; specifically, we varied whether or not the CEO revealed a weakness. To

check the robustness of the results, we varied the location of disclosed weakness across studies: in Study 2A we presented the disclosed weakness at the beginning of the disclosure, and in Study 2B we presented the disclosed weakness at the end of the disclosure.

Pretest. We conducted a pretest to assess whether people deemed the weaknesses we intended to use in Studies 2A and 2B to be sensitive self-disclosures. We presented pretest participants ($N = 302$ U.S. MTurk workers; 159 males; $M_{age} = 36.4$ years, $SD = 11.8$) with a quote from a CEO of a (fictitious) company called RockInvest. There were three between-subjects conditions: In the non-sensitive self-disclosure (i.e., control) condition, participants were told:

> I began my career as a mortgage trader at RockInvest. The company, launched in 1988, initially focused on bonds. But thanks to shrewd acquisitions, the firm is now the world's largest asset manager, with $870 billion, offering a slew of equity funds and multi-asset funds. I take care of my staff, offering health benefits even to part-timers. I like to climb mountains in Colorado and collect American folk art.

In the other two conditions, the above disclosure was followed either with the phrase: "I am nervous about public speaking and I have a habit of cracking my knuckles" or "I am not able to keep track of technological changes." Next, participants were told: "We are interested in your judgement of the sensitivity of information that the CEO disclosed to you above. By 'sensitive' we mean information that is risky for the CEO to disclose, in the sense of making him vulnerable to negative consequences arising from that disclosure." Then, participants responded to the item: "How vulnerable, if at all, is the CEO making himself in disclosing this information?" on a scale ranging from 1 (*not at all*) to 5 (*extremely vulnerable*).

Results confirmed that, relative to the control condition, both weaknesses were perceived as sensitive self-disclosures ($M_{speaking\_weakness}$ = 2.43, SD = 1.11; $M_{control}$ = 2.07, SD = 1.05), $t(208)$ = 2.41, p = .017; ($M_{skills\_weakness}$ = 3.24, SD = 1.10; $M_{control}$ = 2.07, SD = 1.05), $t(193)$ = 7.51, p < .0001. Interestingly, not being able to keep track of technological changes was deemed more sensitive than being nervous about public speaking ($t(197)$ = - 5.23, p < .0001). Study 2A tests the former weakness; Study 2B tests both distinctly.

Study 2A

Study 2A was a two-condition between-subjects design in which the CEO either disclosed (experimental condition) or did not disclose (control condition) a weakness. We predicted that participants would perceive the CEO as more authentic when she disclosed a weakness relative to when she did not and that, in turn, this perceived authenticity would translate into positive overall impressions. We also measured perceived competence and hypothesized that, if disclosing a weakness decreased perceived competence, such a decrement would be more than offset by the authenticity gains of doing so. Thus, we predicted that disclosing a weakness would increase overall impressions and that this effect would be mediated by perceptions of authenticity, even when controlling for perceived competence.

In addition, Study 2A provides evidence of the specificity of our hypothesis surrounding the crucial role of perceived authenticity. Previous research has found that self-disclosure of sensitive information makes a person come across as relatively likeable (Collins & Miller, 1994; Cozby, 1972; Dalto, Ajzen, & Kaplan, 1979; Jourard, 1959; Worthy, Gary, & Kahn, 1969). Therefore, we also measured the extent to which participants viewed the CEO as warm (a construct that encompasses likeability) to ascertain whether the effect of disclosing a weakness on authenticity held when controlling for perceived warmth.

Method

  Procedure. Participants (N = 191 U.S. MTurk workers; 102 males; M_{age} = 34.7 years, SD = 11.2) imagined that during a job interview with the CEO of a (fictitious) company called RockInvest.

  The non-sensitive self-disclosure (i.e., control) condition was the same as in the pretest, in which the CEO did not self-disclose any weaknesses.

  In the experimental condition, the following sentence, in which the CEO self-disclosed flaws, was added to the *beginning* of the description: "I am nervous about public speaking and I have a habit of cracking my knuckles."

  Measures. After participants read the CEO's statement, we assessed four constructs. We assessed perceived authenticity by asking participants to rate the CEO on six dimensions (α = .95: authentic, real, sincere, genuine, inauthentic (reverse-coded), phony (reverse-coded)) on a 7-point scale ranging from 1 (*not at all*) to 7 (*absolutely*), which, when combined, form a measure of perceived authenticity. (The items were adapted from well-established perceived authenticity scales; see Cheshin, Amit, & Van Kleef, 2018; Grandey, Fisk, Mattila, Jansen, & Sideman, 2005; Hahl & Zuckerman, 2014.) Participants provided their overall impressions of the CEO on a scale from -5 (*very unfavorable*) to 5 (*very favorable*). In addition, we assessed competence (three items, α = .92: competent, logical, intelligent) and warmth (four items, α = .93, warm, kind, friendly, easygoing), measured on 7-point scales from 1 (*not at all*) to 7 (*very much*).

Results

  Perceived authenticity. The CEO was perceived as more authentic when he disclosed a weakness relative to when he did not ($M_{experimental}$ = 5.72, SD = 1.09; $M_{control}$ = 5.13, SD = 1.33),

$t(189) = 3.40, p = .0008.$

Overall impressions. Participants had a better impression of the CEO when he disclosed a weakness relative to when he did not ($M_{experimental} = 2.91$, $SD = 1.74$; $M_{control} = 2.30$, $SD = 2.03$), $t(189) = 2.22, p = .028.$

Perceived competence and warmth. The CEO was perceived to be just as competent when he disclosed a weakness relative to when he did not ($M_{experimental} = 5.60$, $SD = 1.17$; $M_{control} = 5.58$, $SD = 1.15$), $t(189) = .15, p = .88$. The CEO was perceived as marginally warmer when he disclosed a weakness relative to when he did not ($M_{experimental} = 5.35$, $SD = 1.25$ vs. $M_{control} = 5.04$, $SD = 1.17$), $t(189) = 1.80, p = .073$. Because our account does not invoke warmth (yet does invoke competence), in the remaining studies, we do not always measure perceived warmth (and when we do so, we do so simply to show that our effects hold when controlling for warmth).

Mediation. A bootstrapping analysis with 10,000 resamples (Hayes, 2017. Model 4) revealed that authenticity mediated the relationship between the presence of sensitive self-disclosure and overall impressions (95% CI: [0.113, 0.388]). This mediation held when controlling for perceptions of competence and warmth (95% CI: [0.093, 0.373]), suggesting that authenticity explains unique variance in the relationship between leaders' sensitive self-disclosures and overall positive impressions.

These results suggest that disclosing foibles makes leaders come across as authentic, leading to enhanced overall impressions.

Study 2B

Study 2B used a hypothetical job-interview paradigm similar to Study 2A, with the following enhancements. First, we added an outcome measure that inquired about a behavior (as opposed to an attitude): We asked participants whether they would be inclined to work for the

given CEO. In addition, in Study 2B, to test for convergent evidence of the effect, we added a

second experimental condition, in which the CEO disclosed a weakness different from that

disclosed in Study 2A – specifically, in this additional condition, the CEO disclosed a skill

weakness ("not being able to keep track of technological changes"). The study was therefore a

three-condition between-subjects design. (In addition, in Supplemental Materials, we report

another conceptual replication, in which the CEO disclosed yet a different weakness, namely: "I

am nervous about teamwork and I have a habit of working alone").

Method

  Procedure. Participants (N = 300 U.S. MTurk workers; 149 males; M_{age} = 35.4 years, *SD*

= 11.6) were randomized to one of three conditions: a *control* condition, in which the CEO did

not disclose a weakness, or one of two experimental conditions, in which she disclosed a

weakness. Specifically, in the *control* condition, participants were told the same description of

the CEO as in Study 2A. In each of the experimental conditions, we added a sentence to the end

of this description. Specifically, in the *speaking weakness* condition, the CEO disclosed: "Even if

I am a manager of a multi-billion dollar company, I am nervous about public speaking and I have

a habit of cracking my knuckles." In the *skills weakness* condition, the CEO disclosed: "Even

though I am a manager of a multi-billion-dollar company, I feel that as the company keeps

growing, I feel a little under the water. The skills the company needs to succeed now are skills I

do not seem to have. I am not able to keep track of the technological changes."

  Measures. We used the same measures of authenticity (α = .92), competence (competent,

logical, intelligent, skillful; α = .91), and overall impressions as in Study 2A. As an additional

measure of the potential positive outcomes that may arise from disclosing weaknesses,

participants were asked whether they would be inclined to accept a job from the given company:

"Suppose you receive a job offer from this company later, how likely are you going to take the job?" using a 7-point scale from 1 (*not at all*) to 7 (*very much*).

Results

      Perceived authenticity. Perceptions of the CEO's authenticity were significantly different by condition ($F(2, 297) = 8.24$, $p = .0003$). Specifically, relative to when she did not disclose weaknesses, the CEO was perceived as more authentic both in the *speaking weakness* condition ($M_{speaking\_weakness} = 4.44$, $SD = .53$; $M_{control} = 4.23$, $SD = .60$), $t(197) = 2.57$, $p = .01$, and in the *skills weakness* condition ($M_{skills\_weakness} = 4.55$, $SD = .56$; $M_{control} = 4.23$, $SD = .60$), $t(199) = 4.01$, $p < .0001$. Perceived authenticity did not differ between the *speaking* and *skills weakness* conditions ($t(198) = 1.42$, $p = .16$).

      Overall impressions. Overall impressions of the CEO differed significantly by condition ($F(2, 297) = 9.31$, $p = .0001$). Specifically, relative to when she did not disclose weaknesses, overall impressions of the CEO were higher in both the *speaking weakness* condition ($M_{speaking\_weakness} = 2.88$, $SD = 1.71$; $M_{control} = 1.98$, $SD = 2.07$), $t(197) = 3.34$, $p = .001$, and the *skills weakness* condition ($M_{skills\_weakness} = 3.04$, $SD = 1.82$; $M_{control} = 1.98$, $SD = 2.07$), $t(199) = 3.85$, $p < .001$. Overall impressions did not differ between the *speaking* and *skills weakness* conditions ($t(198) = .61$, $p = .54$).

      Perceived competence. Perceptions of the CEO's competence did not differ by condition ($F(2, 297) = .71$, $p = .49$).

      Willingness to work. Willingness to work with the leader differed significantly by condition ($F(2, 297) = 4.27$, $p = .015$). Specifically, relative to when the CEO did not disclose weaknesses, participants expressed more interest in working with the CEO in both the *speaking weakness* condition ($M_{speaking\_weakness} = 5.95$, $SD = 1.11$; $M_{control} = 5.43$, $SD = 1.40$), $t(197) = 2.90$,

$p = .004$, and the *skills weakness* condition ($M_{skills\_weakness} = 5.76$, $SD = 1.28$; $M_{control} = 5.43$, $SD = 1.40$), $t(199) = 1.85$, $p = .06$. Willingness to work with the CEO did not differ between the *speaking* and *skills weakness* conditions ($t(198) = 1.04$, $p = .27$).

Mediation. Comparing the *speaking weakness* condition to the *control* condition, bootstrapping analyses (with 10,000 resamples) showed that authenticity mediated the relationship between weakness disclosure and overall impressions, controlling for perceptions of competence: the index of mediation excluded zero (95% CI = [0.073, 0.726]), suggesting a significant indirect effect (Hayes, 2017). We observed similar results when we used willingness to work with the leader as the dependent variable (95% CI = [0.024, 0.371]). Similarly, comparing the *skills weakness* condition to the *control* condition, bootstrapping analyses (with 10,000 resamples) showed that authenticity mediated the relationship between weakness disclosure and overall impressions, controlling for perceptions of competence: the index of mediation excluded zero (95% CI = [.092, .467]), suggesting a significant indirect effect (Hayes, 2017). We observed similar results when we used willingness to work with the leader as the dependent variable (95% CI = [0.049, 0.268]).

In sum, Studies 2A and 2B provide preliminary evidence that when leaders disclose their foibles, it makes them come across as authentic (Studies 2A & 2B), culminating in positive outcomes, such as making a positive impression on followers (Study 2A) and increasing followers' desire to work with that leader (Study 2B).

Our effects are robust to whether the weakness is presented (at the beginning of the disclosure, as in Study 2A) versus at the end of the passage (as in Study 2B). This robustness is noteworthy, as it distinguishes our effect from a related phenomenon: incorporating a small dose of negative information in product descriptions can lead to positive evaluations (Ein-Gar, Shiv,

& Tormala, 2011). Ein-Gar et al. (2011) show that this effect arises because negative

information, when placed after positive information, makes the positive information more

salient. Their effect does not emerge when the negative information is presented first. By

contrast, our effect holds regardless of whether the weakness is disclosed upfront versus prefaced

with the disclosure of neutral or desirable information.

In addition, our effects are robust to the leader's gender. In Studies 2A and 2B (and all

subsequent scenario studies), we did not specify the gender of the disclosing leader. However, in

the Supplemental Materials, we report an experiment in which we manipulated the gender of the

disclosing leader (in addition to whether the leader revealed a weakness or not) and found only a

main effect of revealing a weakness.

Study 3: Behavioral Outcome

Whereas in Studies 2A and 2B, participants simply read about a CEO's disclosure, in

Study 3, we induced participants to make a sensitive disclosure (or not) to other participants.

Specifically, Study 3 was a simulated employment task in which participants were randomized to

the role of either manager or prospective employee; managers were further randomized to either

disclose or not disclose a weakness to that employee participant. The study was therefore a 2x2

between-subjects design manipulating role (manager versus employee) and disclosure content

(weakness disclosed versus no weakness disclosed).

Prospective employees rated their perceptions of the manager's authenticity, competence,

as well as their desire to work with that manager. And, as a behavioral measure of the potential

benefits of managers' sensitive self-disclosure, prospective employees engaged in a trust game

(Berg, Dickhaut, & McCabe, 1995) with the manager; we measured how much, if any, money

prospective employees decided to transfer to the manager (in so doing, making themselves

vulnerable; for, per the standard trust game, managers could keep all of the money for themselves or choose to send a portion back to the employee). We predicted that prospective employees would be more trusting of managers who made sensitive self-disclosures relative to those who did not.

In addition, Study 3 began to assess whether would-be disclosers are aware of the benefits of sensitive self-disclosure that we have been documenting (a line of inquiry that we continue in Study 7). Specifically, in Study 3, we asked managers to predict how they thought their employees would perceive them. We wondered whether managers induced to disclose a weakness would accurately predict that doing so would boost their overall impression (relative to managers randomized to *not* disclose a weakness).

Method

Procedure. Participants ($N = 218$; 99 males; $M_{age} = 22.3$ years, $SD = 4.3$) came to a lab at a northeastern university and received a $15 base payment plus study earnings, as described below.

In a simulated hiring task, we randomized half of participants to the role of manager and the other half to the role of prospective employee, and randomly grouped participants into manager-prospective employee dyads. Participants began the session in individual cubicles, where they were informed of their assigned role and given information on the task to follow.

At the start of the study, prospective employees were told that they would be participating in a simulated employment task and that in a moment, they would meet their potential manager, who would have a task for them to complete. Therefore, they were told, the manager would be evaluating the prospective employee's performance on the task.

Managers were informed that they would be meeting their potential employee for a position on their team and would assign the employee a ten-item "word correction" task (see Supplemental Materials). Managers were told that they would time the employee on the task for one minute. In reality, we used the word-correcting task as a cover story and it was not graded. Managers were further informed that in a few minutes, they would meet their employee, at which point they should introduce themselves using a script provided for this purpose. Critically, the script manipulated whether managers would disclose a weakness. Specifically, in the control condition, managers were instructed to introduce themselves by saying:

Hi, I am [*name*], the manager of the organization. I am going to direct the task and the standards by which the work is to be evaluated. In addition, I will also evaluate you at the end of the session in a private questionnaire. Let me introduce myself a little bit: I am the president of the graduate student association at the university. I get to travel often to cities across the country to give presentations. I enjoy what I do.

In the experimental condition, the script was the same, except that the following sentences were appended to the end: "I'm quite shy. I am nervous about public speaking, and I have a habit of cracking my knuckles." Next, participants were randomly assigned to prospective employer-manager dyads; each dyad was ushered into their own private room to complete the task Managers were given a few minutes to practice the script so that they could deliver it from memory, without a written script, when they introduced themselves to the employee. Next, the managers assigned the employee the task and, using a stopwatch, gave them one minute to complete it.

Survey measures. After the task, participants returned to their individual cubicles. Prospective employees provided their assessments of their managers' authenticity ($\alpha = .88$), and

competence ($\alpha = .88$) as in Studies 2A and 2B, as well as their desire to work with the manager:

"Would you want to be paired with this manager again for a subsequent task?" measured on 7-point scales from 1 (*want to work with a different manager*) to 7 (*want to work with this manager*).

Managers indicated how they thought the prospective employee viewed them; specifically, we asked managers: "What do you think is the prospective employee's overall impression of you as a manager?" on a scale ranging from -5 (*very unfavorable*) to 5 (*very favorable*), and: "Do you think the prospective employee would want to be paired with you as a manager again for a subsequent task?" on a scale ranging from 1 (*want to work with a different manager*) to 7 (*want to work with this manager*). In addition, we asked managers: "What is your overall impression of the prospective employee" on a scale ranging from -5 (*very unfavorable*) to 5 (*very favorable*) (administered because we thought managers would find it odd if we did not ask this, given that the task was a job-interview simulation). Finally, both managers and prospective employees answered whether they knew each other before the experiment. Three pairs of dyads knew each other previously and therefore were excluded from the data analysis, leaving 212 people. The results are substantively equivalent when these six individuals are included in the analysis.

Behavioral outcome measure. We assessed employees' trust in the manager via a trust game; employees were assigned the role of sender, and managers were assigned the role of receiver. First, we explained the game to all participants, telling them that the employee will receive an initial endowment of $3 in quarters ($0.25), and will have to decide how much of this amount to transfer to the manager. Any amount transferred will be tripled. Then the manager has

to decide how much of this tripled amount s/he would like to send back to the employee. The amount sent back will not be tripled.

Participants were encouraged to ask questions or re-read instructions if they did not understand how the game worked. Upon checking a box labelled "I understand how the game works," participants proceeded to the trust game, with each employee indicating how much, if any, money to transfer to their manager, and with the manager then indicating how much, if any, of this money to return to their employee. Participants were given real money to play the trust game. At the end of the experiment, participants provided demographic information and were debriefed.

Results

Employees

Perceived authenticity. Prospective employees perceived their manager as more authentic when that manager disclosed a weakness relative to when s/he did not ($M_{weakness\_disclosed}$= 5.39, SD = 1.10; $M_{no\_weakness\_disclosed}$ = 4.87, SD = 1.09, $t(104)$ = 2.46, p < .02).

Overall impression. Prospective employees had a better impression of their manager when the manager disclosed a weakness relative to when s/he did not ($M_{weakness\_disclosed}$= 3.00, SD = 1.50; $M_{no\_weakness\_disclosed}$ = 2.43, SD = 2.00, $t(104)$ = 2.75, p < .10).

Perceived competence. Prospective employees' perceptions of their manager's competence did not differ by condition ($M_{weakness\_disclosed}$ = 5.18, SD = 1.02; $M_{no\_weakness\_disclosed}$ = 5.10, SD = 1.17, $t(104)$ = .37, p = .71).

Willingness to work with the manager again. Prospective employees were more interested in continuing to work with their manager when that manager disclosed a weakness relative to when s/he did not ($M_{weakness\_disclosed}$= 5.41, SD = 1.34; $M_{no\_weakness\_disclosed}$ = 4.78, SD =

1.38), $t(104) = 2.37$, $p < .02$).

Trust. As an indicator of trust, prospective employees transferred more money to the manager when that manager disclosed a weakness relative to when s/he did not ($M_{weakness\_disclosed}$ = $2.39, $SD = .84$; $M_{no_weakness_disclosed} = 2.00, $SD = 1.07$, $t(104) = 2.04$, $p = .043$).

Mediation. We conducted a mediation analysis with overall impressions as the dependent variable to test mediation by authenticity while controlling for competence. A 10,000-sample bootstrap analysis (Hayes, 2017, Model 4) showed that the index of mediation excluded zero (95% CI: [0.075, 0.604]), suggesting a significant indirect effect. We observed similar mediation analyses results with trust (i.e., money allocation, 95% CI: [0.042, 0.366]) and willingness to work with the manager again (95% CI: [0.065, 0.500]) as dependent variables, suggesting significant indirect effects.

Managers. There was no difference in managers' overall impressions of the paired employees ($M_{weakness\_disclosed} = 2.11$, $SD = 1.71$; $M_{no\_weakness\_disclosed} = 2.02$, $SD = 1.86$, $t(104)$ = .26, $p = .79$). Managers predicted that their employees would view them similarly, regardless of whether they disclosed their weaknesses. Specifically, there were no differences between conditions in managers' predictions of the employee's overall impressions of them ($M_{weakness\_disclosed} = 2.10$, $SD = 1.33$; $M_{no\_weakness\_disclosed} = 1.76$, $SD = 1.59$, $t(104) = 1.18$, $p = .24$). Similarly, there were no differences between conditions in managers' predictions of whether the prospective employee would want to be paired with them again for a subsequent task ($M_{weakness\_disclosed} = 4.81$, $SD = 1.03$; $M_{no\_weakness\_disclosed} = 4.56$, $SD = 1.19$, $t(104) = 1.16$, $p = .25$).

In sum, Study 3 provides converging evidence that sensitive self-disclosure makes leaders come across as authentic, leading to positive outcomes – here, operationalized as employees' increased trust in, and desire to work with, that leader. In addition, Study 3 provides

preliminary evidence that would-be disclosers are unaware of these benefits: managers induced

to make a sensitive self-disclosure did not appear to anticipate that doing so would cause their

employees to want to work with them.

Study 4: Disclosure Sensitivity

So far, we have shown that when leaders reveal their weaknesses, followers perceive

them as more authentic, which can produce positive outcomes. However, this is not to say that

revealing *any* weakness, no matter how great, will produce such effects. Indeed, after all, there

are risks to sensitive self-disclosure in general and to disclosing weaknesses in particular. Thus,

in Studies 4, 5, and 6, we test boundaries of our effects. We begin by manipulating the severity

of the weakness and measuring the point at which the authenticity gains from making such

disclosures are more than offset by decrements in perceived competence. Specifically, we test

our "sensitivity sweet spot" hypothesis, such that mild to moderately sensitive disclosures – i.e.,

a leader's disclosure of her legitimate but not damning weaknesses – produce positive outcomes.

Study 4 used the same basic paradigm as Studies 2A and 2B, in which participants

imagined what a CEO told them during a job interview. Between subjects, we manipulated the

severity of the weakness that the CEO disclosed – five different levels, ranging from mildly

sensitive to exceptionally sensitive. The mildly sensitive weakness was similar to that used in the

experimental conditions so far, and thus is a level that we know to reliably produce positive

impressions relative to not revealing a weakness. We predicted that the mildly sensitive

weakness would produce better overall impressions relative to the exceptionally sensitive

condition, where we posited the competence decrement would offset any authenticity gains. In

this vein, we predicted that at some point between these extremes, the overall positive impression

afforded by disclosing weaknesses would "tip" toward being negative, though we did not predict

the specific point at which that would happen. This is why we tested five different severity

levels.

Study 4 also provides additional evidence of the specificity of our basic effect – namely,

that the positive outcomes arising from leaders' self-disclosure of weaknesses is driven by

enhanced perceptions of their authenticity, as opposed to, for example, decreased perceived

status. For example, perhaps revealing weaknesses leads followers to view the leader as more

"on their level" status-wise, making them more "accessible" or "relatable," which could make

for an overall positive impression. Indeed, prior work has found that revealing vulnerabilities can

decrease perceived status (Gibson et al., 2018).

Method

Procedure. As in Studies 2A and 2B, participants ($N = 754$ U.S. MTurk workers; 54%

male*; M*$_{age}$ = 35.5 years, $SD = 10.7$) read a statement that a CEO ostensibly had made. Between

subjects, we manipulated the sensitivity of the disclosure. Specifically, the statement "I take care

of my staff, offering health benefits even to part-timers. I like to climb mountains and collect

American folk art" was appended with one of five different disclosures of weaknesses, varying

in sensitivity, as follows:

Mildly sensitive: I am nervous about public speaking.

Sensitive: I am very nervous about public speaking.

Moderately sensitive: I am very nervous about public speaking: when I speak to groups larger

than two people, my mouth gets dry.

Extremely sensitive: I am very nervous about public speaking: when I speak to groups larger than

two people, my mouth gets dry and I sometimes start to panic.

Exceptionally sensitive: I am very nervous about public speaking: when I speak to groups larger than two people, my mouth gets dry and I sometimes start to panic. One time I was giving a speech to investors and had a panic attack.

Measures. As in Study 2A, we measured perceived authenticity (α = .90), competence (α = .84), warmth (α = .86), and overall impression (-5 = *very unfavorable* to 5 = *very favorable*). We also included two measures of status: (1) a self-reported measure ("I admire the CEO," "I respect the CEO," "I think the CEO is of high-status," 7-point scale from 1 (*strongly disagree*) to 7 (*strongly agree*), α = .85), and (2) a status ladder measure: "Below you will see a ladder with 9 rungs. Think of this ladder as representing where people stand in a group. At the top of the ladder are the people who are most respected, admired, and held in highest regard. At the bottom are the people who are the worst off, those who are least respected, admired, and held in lowest regard. Where would you place this person on this ladder after you interact with him? Please drag the slider and answer the question" (9-point scale, Magee & Galinsky, 2008).

Results

Analyses revealed significant differences among conditions on authenticity ($F(4, 749) = 3.94$, $p = .004$), competence ($F(4, 749) = 3.10$, $p = .01$), warmth ($F(4, 749) = 4.86$, $p = .0007$), and overall impressions ($F(4, 749) = 4.34$, $p < .002$), but no significant difference on any of the status measures: self-reported status ($F(4, 749) = 1.49$, $p = .20$) or the status-ladder measure ($F(4, 749) = 1.93$, $p = .103$). Table 2 presents all means and standard deviations by condition.

As predicted, there was a "tipping point" in terms of disclosure sensitivity. Specifically, when the weakness was mild, increasing severity – i.e., moving from mildly sensitive to sensitive – boosted overall impressions ($M_{sensitive} = 3.45$ vs. $M_{mildly\ sensitive} = 3.03$, $t(749) = 2.15$, $p < .04$). However, beyond sensitive disclosure, increasing sensitivity hurt overall impressions – that is,

relative to the sensitive condition, overall impressions were significantly lower in the moderately

sensitive condition ($M_{\text{moderately sensitive}}$ = 2.75 vs. $M_{\text{sensitive}}$ = 3.45, $t(749)$ = -3.49, p = .0005), the

extremely sensitive condition ($M_{\text{extremely sensitive}}$ = 2.78 vs. $M_{\text{sensitive}}$ = 3.45, $t(749)$ = -3.31, p

= .001), and the exceptionally sensitive condition ($M_{\text{exceptionally sensitive}}$ = 2.79 vs. $M_{\text{sensitive}}$ = 3.45,

$t(749)$ = -3.34, p = .0009). Overall impressions in the three most sensitive disclosure conditions

were equivalent ($M_{\text{moderately sensitive}}$ = 2.75 vs. $M_{\text{extremely sensitive}}$ = 2.78 vs. $M_{\text{exceptionally sensitive}}$ = 2.79, p

= n.s.).

Mediation. Bootstrapping analyses (with 10,000 resamples) showed that authenticity

mediated the relationship between weakness disclosure (treated as a continuous variable) and

overall impressions (95% CI: [0.023, 0.139]). Importantly, this mediation held when controlling

for perceptions of competence and warmth (95% CI: [0.048, 0.111]). Status did not mediate the

relationship between weakness disclosure and overall impressions (95% CI: [- 0.097, 0.037]).

In sum, Study 4 suggests that when it comes to making a good impression, it can behoove

leaders to reveal weaknesses. However, these benefits were restricted to mild-to-moderately-

sensitive disclosures. Beyond this point, the authenticity gains of disclosing one's weaknesses

were more than offset by decrements in perceived competence, worsening one's overall

impressions.

In addition, Study 4 provides additional evidence attesting to the specificity of our effect,

namely that the positive impression left by leaders who reveal mild-to-moderate weaknesses is

driven by perceived authenticity; such disclosures did *not* impact status perceptions. This result

may appear incongruent with that of recent work by Gibson et al. (2018), who showed that in an

interactive task, when a person discloses vulnerability, it can cause their partner to view them as

lower status, reducing their influence on their partner. One possibility for the difference in

findings is the clarity of status. In our experiment, the manager was clearly – by definition, in

fact – higher status than the prospective employee (i.e., respected and admired, Magee &

Galinsky 2008). In Gibson et al.'s fascinating study, the discloser was an MBA student, and the

partner was an undergraduate student, and it is not entirely clear whether undergraduates

perceived MBA students as higher status (or whether MBA students perceived undergraduates as

lower status). Relative to the present simulated employment task, where the status hierarchy was

made abundantly clear, there was greater status ambiguity in Gibson et al.'s study. We suggest

that disclosing one's vulnerabilities may be particularly likely to decrease status in situations in

which there is status ambiguity. In this vein, in Study 5, we test whether self-disclosing a mild-

to-moderate weakness only produces beneficial outcomes when the discloser is high status.

Study 5: Status

In Study 5, we test the hypothesis that positive outcomes arising from self-disclosing

weaknesses are restricted to high-status disclosers. The study was a 2x2 between-subjects design

in which we manipulated the discloser's status (high versus low) as well as the disclosure content

(weakness disclosed vs. no weakness disclosed).

We predicted that participants would form a positive impression of the high-status, but

not low-status, colleague who disclosed a weakness. Consistent with previous studies, we also

predicted that this increased positive impression would arise via gains in perceived authenticity.

Thus, we predicted moderated mediation: the capacity for sensitive self-disclosure to foster

perceptions of authenticity, in turn boosting overall impression, would be restricted to the high-

status disclosers.

Method

Procedure. Participants ($N = 183$; 97 males; $M_{age} = 31.5$ years, $SD = 12.0$) from a city in

the northeastern United States participated in this study for pay. This study was part of an hour-

long series of studies for which participants received $20 as compensation. They were asked to

imagine that they were prospective employees engaged in a job interview with a company

employee named Elis. We manipulated discloser status by telling participants to imagine that

they were either meeting with the CEO (high-status condition) or a junior analyst (low-status

condition).

All participants were asked to imagine that their interviewer said:

The company, launched in 1988, initially focused on bonds. But thanks to shrewd

acquisitions, the firm is now the world's largest asset manager, with $870 billion, offering

a slew of equity funds and multi-asset funds. I like to climb mountains and collect

American folk art in my spare time.

In the no-weakness-disclosed condition, the following sentence was appended to the

above script: "I need to fly several times a month, and I enjoy flying." In the weakness-disclosed

condition, the final sentence was: "Even though I need to fly several times a month, I am actually

afraid of flying and I avoid flying."

Pretest. We conducted a pretest to ascertain whether the disclosed weakness was

perceived as similarly sensitive across the status manipulation. Specifically, we randomly

assigned participants ($N = 391$ U.S. MTurk workers; 172 males; $M_{age} = 37.0$ years, $SD = 11.8$) to

one of the above four conditions in a 2 (weakness vs. no-weakness) x 2 (high vs. low status)

between-subjects design. We told these pretest participants: "We are interested in your judgment

of the sensitivity of information that Elis disclosed to you above. By 'sensitive' we mean

information that is risky for Elis to disclose, in the sense of making him vulnerable to negative

consequences arising from that disclosure." We then asked them to indicate "How vulnerable, if

at all, is Elis making himself in disclosing this information?" on a scale from 1 to 5 (1 = *not at all*; 5 = *extremely vulnerable*). Results of a 2x2 ANOVA revealed only a main effect of disclosing a weakness ($F(1, 387) = 18.31, p < .0001$), such that disclosing "afraid of flying" made Elis more vulnerable relative to when s/he did not do so ($M_{Weakness}= 2.15, SD = .85; M_{No-Weakness} = 1.79, SD = .81, t(387) = 4.38, p < .0001$). Importantly, no other main effect or two-way interaction was significant, which indicates that the weakness disclosure was seen as equally sensitive regardless of status and, hence, the disclosure manipulation was equally strong as a function of status.

Measures. Participants responded to the same items used in previous studies: perceptions of authenticity ($\alpha = .93$), competence ($\alpha = .89$), overall impression, and willingness to work with the person from the company.

Results

Perceived authenticity. A 2x2 ANOVA revealed only an interaction ($F(1, 179) = 4.59, p = .033$). Specifically, disclosing a weakness boosted perceived authenticity of the high-status discloser ($M_{Weakness\_HighStatus}= 5.01, SD = 1.16; M_{NoWeakness\_HighStatus} = 4.28, SD = 1.50, t(76) = 2.34, p = .02$) but not of the low-status discloser ($M_{Weakness\_LowStatus} = 4.48, SD = 1.46; M_{NoWeakness\_LowStatus} = 4.63, SD = 1.34, t(103) = 1.34, p = .58$).

Perceived competence. A 2x2 ANOVA revealed a main effect of status ($F(1, 179) = 3.87, p = .051$), such that perceived competence was (marginally) greater for high-status than low-status disclosers ($M_{HighStatus} = 5.70, SD = 1.16; M_{LowStatus} = 5.37, SD = 1.11; t(179) = 1.97, p = .051$), and an interaction ($F(1, 179) = 5.48, p = .02$), such that disclosing a weakness did not hurt competence perceptions of high-status disclosers ($M_{Weakness\_HighStatus} = 5.89, SD = .92; M_{NoWeakness\_HighStatus} = 5.50, SD = 1.33, t(76) = 1.30, p = .20$) but marginally decreased

competence perceptions of low-status disclosers ($M_{Weakness\_LowStatus}$ = 5.17, SD = 1.24;

$M_{NoWeakness\_LowStatus}$ = 5.57, SD = .91, $t(103)$ = 1.83, p = .070).

Overall impressions. Mirroring the results of the authenticity measure, a 2x2 ANOVA

revealed a marginally significant interaction ($F(1, 179)$ = 3.38, p = .068), such that disclosing a

weakness boosted overall impression of the high-status discloser ($M_{Weakness\_HighStatus}$ = 2.19, SD =

2.00; $M_{NoWeakness\_HighStatus}$ = 1.05, SD = 2.54, $t(76)$ = 2.22, p < .03) but not of the low-status

discloser ($M_{Weakness\_LowStatus}$ = 1.25, SD = 2.34 vs. $M_{NoWeakness\_LowStatus}$ = 1.36, SD = 2.11, $t(103)$

= .24, p = .81).

Willingness to work with the discloser. Mirroring the results of both the authenticity

and overall impressions measures, a 2x2 ANOVA revealed a significant interaction ($F(1, 179)$ =

4.01, p = .047), such that disclosing a weakness increased willingness to work with the discloser

when interacting with a high-status person who disclosed a weakness ($M_{Weakness\_HighStatus}$ = 5.35,

SD = 1.57; $M_{NoWeakness\_HighStatus}$ = 4.61, SD = 2.02, $t(76)$ = 1.98, p = .049) and not when

interacting with a low-status person who disclosed a weakness ($M_{Weakness\_LowStatus}$ = 4.87, SD =

1.36; $M_{NoWeakness\_LowStatus}$ = 5.12, SD = 1.66, $t(103)$ = .77, p = .44) (Figure 2).

Moderated mediation. We conducted a moderated-mediation analysis with overall

impressions as the dependent variable to simultaneously test moderation by status and mediation

by perceptions of authenticity. A 10,000-sample bootstrap analysis (Model 7) showed that the

index of moderated mediation excluded zero (95% CI = [-2.154, -.096]), suggesting a significant

indirect effect (Hayes, 2017). We also conducted the same moderated mediation analysis with

willingness to work with the discloser as the dependent variable. A 10,000-sample bootstrap

analysis (Model 7) showed that the index of moderated mediation excluded zero (95% CI = [-

1.071, -.055]), suggesting a significant indirect effect (Hayes, 2017).

In sum, Study 5 suggests that for disclosing foibles to culminate in overall positive impressions, the disclosure must be made by high-status rather than low-status individuals.

Study 6: Voluntariness

In Study 6, we test the hypothesis that for leaders to reap benefits from self-disclosing their foibles, the disclosure must be made of their own volition, as opposed to by requirement. The study was a 2x2 between-subjects design in which we manipulated the content of the disclosure (weakness disclosed vs. no weakness disclosed), as well as the voluntariness of the disclosure (voluntary vs. required).

We predicted an interaction such that a leader's sensitive self-disclosure would foster positive impressions only when that leader did so voluntarily, as opposed to by requirement. Consistent with the previous studies, we also predicted that this increased positive impression would arise via gains in perceived authenticity. Thus, again, as in Study 5, we predicted moderated mediation: the capacity for sensitive self-disclosure to foster perceptions of authenticity, in turn boosting overall impression, would be restricted to the situation in which the leader disclosed voluntarily.

Method

Procedure. Participants ($N = 392$ U.S. MTurk workers; 174 males*; $M_{age} = 29.8$ years, *SD* $= 12.3$) read how a previous participant had ostensibly introduced themselves in a prior experiment:

> I am a manager of a technological company. I began my career as an engineer at this
>
> company. Thanks to shrewd acquisitions, the firm is now one of the big companies in the
>
> field. As a manager, I take care of my staff, offering health benefits even to part-timers. I

am good at leading detail-oriented projects. I have won numerous awards in my

field. I like to climb mountains in Colorado and collect American folk art.

We manipulated whether the manager disclosed a weakness by, for half of participants,

appending the following self-disclosure of a weakness: "Even though I am a manager of the

company, I am nervous about public speaking and I have a habit of cracking my knuckles."

We manipulated whether the manager's disclosures were made voluntarily by informing

half of participants that the disclosure had been required. (In the voluntary condition, we simply

omitted this note, on the assumption that, unless stated otherwise, participants would assume that

the disclosure had been made voluntarily.) This notice was necessarily different depending on

whether the manager had disclosed a weakness. Specifically, when the manager had disclosed a

weakness, participants randomized to the required disclosure condition were further told that: "In

the previous study, the individual was required to include negative self-relevant information in

the introduction." Similarly, when the manager had not disclosed a weakness, participants

randomized to the required disclosure condition were told that "In the previous study, the

individual was required to include only positive self-relevant information in the introduction."

Measures. Participants responded to the same items measuring the perceptions of

authenticity ($\alpha = .92$) and competence ($\alpha = .88$) of the manager, overall impressions, and

willingness to work with the manager as in Study 2B.

Results

Perceived authenticity. A 2x2 ANOVA revealed a main effect of disclosing a weakness

($F(1, 388) = 17.43$, $p < .0001$), such that the manager was viewed as significantly more authentic

when s/he disclosed a weakness relative to when s/he did not ($M_{weakness\_disclosed} = 5.57$, $SD = 1.04$;

$M_{only\_strengths\_disclosed} = 5.11$, $SD = 1.06$, $t(388) = 4.17$, $p < .0001$). However, this main effect was

qualified by a significant interaction ($F(1, 388) = 4.90$, $p = .027$), suggesting that, if anything,

the authenticity benefit of revealing a weakness was driven by managers who had done so

voluntarily, as opposed to by requirement. Specifically, when managers voluntarily revealed a

weakness, they were perceived as more authentic relative to when they only voluntarily revealed

their strengths ($M_{weakness\_volunteered} = 5.75$, $SD = .96$; $M_{only\_strengths\_volunteered} = 5.06$, $SD = 1.06$, $t(193)$

$= 4.52$, $p < .0001$). However, when managers were required to reveal a weakness, this difference

disappeared ($M_{weakness\_required} = 5.39$, $SD = 1.17$; $M_{only\_strengths\_required} = 5.18$, $SD = 1.02$, $t(195) =$

1.39, $p = .17$).

Perceived competence. A 2x2 ANOVA revealed only a marginally significant main

effect of disclosing a weakness ($F(1, 388) = 3.43$, $p = .065$), such that the manager was viewed

as marginally more competent when s/he disclosed a weakness relative to when s/he did not do

so ($M_{weakness\_disclosed} = 5.52$, $SD = .92$, $M_{only\_strengths\_disclosed} = 5.36$, $SD = .91$).

Overall impressions. Mirroring the results of the authenticity measure, a 2x2 ANOVA

revealed a main effect of disclosing a weakness ($F(1, 388) = 6.35$, $p = .012$), such that the

manager left a better overall impression when s/he disclosed a weakness relative to when s/he

did not do so ($M_{weakness\_disclosed} = 2.75$, $SD = 1.67$; $M_{only\_strengths\_disclosed} = 2.32$, $SD = 1.76$, $t(388) =$

2.52, $p = .012$). And, as with the authenticity measure, this main effect was qualified by an

interaction ($F(1, 388) = 5.49$, $p < .02$), such that disclosing a weakness only boosted overall

impressions when done voluntarily ($M_{weakness\_volunteered} = 2.98$, $SD = 1.60$; $M_{only\_strengths\_volunteered} =$

2.14, $SD = 1.73$, $t(193) = 3.44$, $p = .0006$), as opposed to by requirement ($M_{weakness\_required} = 2.53$,

$SD = 1.70$; $M_{only\_strengths\_required} = 2.50$, $SD = 1.72$, $t(195) = .12$, $p = .90$).

Willingness to work with the leader. A 2x2 ANOVA revealed only a marginally

significant interaction ($F(1, 388) = 3.43$, $p = .064$), such that disclosing a weakness only

increased willingness to work with the manager when s/he disclosed a weakness when done voluntarily ($M_{weakness\_volunteered} = 5.85$, $SD = 1.04$; $M_{only\_strengths\_volunteered} = 5.51$, $SD = 1.11$, $t(193) = 2.19$, $p < .03$), as opposed to by requirement ($M_{weakness\_required} = 5.59$, $SD = 1.00$; $M_{only\_strengths\_required} = 5.66$, $SD = 1.15$, $t(195) = -.43$, $p = .66$) (Figure 3).

Moderated mediation. We conducted a moderated mediation analysis with overall impressions as the dependent variable to simultaneously test moderation by voluntariness and mediation by perceptions of authenticity, controlling for perceptions of competence. A 10,000-sample bootstrap analysis (Hayes, 2017, Model 7) showed that the index of moderated mediation excluded zero (95% CI = [.006,.613]), suggesting a significant indirect effect. A similar moderated mediation with willingness to work with the leader (10,000 sample bootstrap analysis, Hayes, 2017, Model 7) showed that the index of moderated mediation also excluded zero (95% CI = [.004,.261]), suggesting a significant indirect effect.

In sum, Study 6 suggests that for leaders to reap benefits from self-disclosing their foibles, the disclosure must be voluntary as opposed to required. This finding also squarely distinguishes our effect from the pratfall effect, whereby the knowledge that someone has made a blunder, such as spilling a cup of coffee, can increase likeability (Aronson, Willerman, & Floyd, 1966). Unlike the present effect, the pratfall effect does not hinge on a person's (voluntary) self-disclosure of a weakness.

Study 7: Disclosers' Intuitions

Study 3 provided preliminary evidence that would-be disclosers may be unaware of the benefits of disclosing weaknesses that we have documented thus far. In Study 7, we provide further evidence that would-be disclosers do not intuit these benefits. Study 7 uses the same paradigm as Study 2B, with the exception that instead of asking participants to take the

perspective of the prospective employee receiving an interviewer's disclosures, we asked them to imagine that they were the company leader who either had or had not disclosed a weakness and to predict how they would come across to the prospective employee. The study was therefore a two-condition between-subjects design. Importantly, we used a weakness from a prior study (Study 5) – i.e., a weakness that we have already shown to foster increased overall positive impressions in observers.

Method

    Procedure. Participants ($N = 140$; 71 males, $M_{age} = 29.2$ years, $SD = 11.1$) came to a lab at a northeastern university. The study was part of an hour-long series of studies for which participants received \$20 as compensation. Participants imagined that they were a leader at a company and engaged in self-disclosure to a job candidate during an interview. Specifically, they were asked to imagine that they had said the following to the candidate:

> When I joined the company, the company was in deep trouble and on the brink of terrible economic times. I play hard on negotiation. I extracted \$2 billion from another marketing company to dissolve a partnership with our company, securing the funding to jump start a company mired in years of losses. I cut costs and sharpened sales practices to gain market shares and boost margins. Sales have kept climbing after that. I like to climb mountains and collect American folk art in my spare time.

    In the control condition, the following sentence was appended to the above text: "I need to fly several times a month, and I enjoy flying." In the experimental condition, in which we wanted participants to suppose that they had made a sensitive self-disclosure, the following sentence was appended to the above text: "Even though I need to fly several times a month, I am actually afraid of flying and I avoid flying" – i.e., the same weakness used in Study 5.

Measures. The measures were similar to those in our previous studies, except from the

perspective of the discloser. In other words, we asked participants how they thought the job

candidate would rate their authenticity ($\alpha = .95$), competence ($\alpha = .90$), and overall impression,

as well as to predict the candidate's willingness to work with them.

Results

Predicted perceptions of authenticity. Participants thought they would come across as

similarly authentic, regardless of whether they disclosed a weakness ($M_{experimental} = 4.76$, $SD =$

1.25; $M_{control} = 4.61$, $SD = 1.54$, $t(138) = .64$, $p = .523$).

Predicted perceptions of competence. Participants thought they would come across as

less competent when they disclosed a weakness ($M_{experimental} = 5.05$, $SD = 1.29$; $M_{control} = 5.78$,

$SD = 0.95$, $t(138) = 3.82$, $p < .002$).

Predicted overall impressions. Mirroring the results for predicted authenticity

perceptions, participants thought they would make a similar overall impression, regardless of

whether they disclosed a weakness ($M_{experimental} = 1.55$, $SD = 2.18$; $M_{control} = 2.13$, $SD = 2.18$,

$t(138) = 1.57$, $p = .12$).

Predicted propensity to work with the leader. Mirroring the results for authenticity

perceptions and overall impressions, participants thought that the candidate's willingness to work

with them would be similar, regardless of whether they disclosed a weakness $M_{experimental} = 5.15$,

$SD = 1.04$; $M_{control} = 5.33$, $SD = 1.22$, $t(138) = .93$, $p = .353$).

In sum, consistent with Study 3, Study 7 suggests that would-be disclosers do not predict

the positive effects of disclosing a weakness that we have shown in Studies 1-6, suggesting that

leaders may under-employ a tool for creating a positive impression: revealing their foibles.

General Discussion

Although authenticity in organizations has benefits, leaders, in particular, face barriers to being perceived as such (Hahl & Zuckerman, 2014). We show that leaders can increase perceptions of their authenticity by engaging in mild-to-moderately sensitive self-disclosures. Beyond this point, the authenticity gains that leaders get from disclosing weakness appear to be more than offset by decrements in perceived competence, worsening overall impressions. Also consistent with our conceptual account, we documented that the self-disclosure needs to be made voluntarily from a high-status individual (in this investigation, leaders and managers).

Contribution to Theory and Practice

Our research makes contributions to the self-disclosure literature. Whereas past research has emphasized the relationship between self-disclosure and liking (Collins & Miller, 1994), we focus on the role of self-disclosure in work relationships, specifically in the context of leader-follower relationships. Similar to Gibson et al. (2018), our work broadens the scope of self-disclosure from dyad relationships to organizational-relevant settings. Complementing Gibson et al. (2018), our work suggests that perceived authenticity is a unique input to work-relevant interpersonal outcomes in addition to perceived status and that there exists a "sweet spot" wherein a leader can boost followers' perceptions of their authenticity without suffering a decrement in perceived status.

Our research contributes to the leadership literature by providing guidance on how leaders can enhance perceived authenticity. Given that leaders are prone to being perceived as inauthentic (Hahl & Zuckerman, 2014) and given the organization-level benefits of having authentic leaders (Avolio et al., 2004; Norman et al., 2010), uncovering how, when, and whose self-disclosure can boost perceived authenticity is important. We suggest that for leaders to

realize the benefits of sensitive self-disclosure, the disclosure has to be voluntary and mild to moderately sensitive in nature.

Finally, our research contributes to the self-presentation literature by uncovering one way to soften the "braggart" image that is associated with self-promotion. Motivated by self-presentation concerns, actors seek to maximize their perceived competence by self-promoting; however, self-promotion can decrease liking without boosting perceived competence (Scopelliti, Loewenstein, & Vosgerau, 2015). We suggest that by disclosing mild to moderately sensitive foibles, leaders may be able to come across as more authentic and generate more favorable outcomes without diminishing perceptions of their competence.

Limitations and Future Research Directions

One important future research direction would be to examine whether self-disclosure of weaknesses works equally for different people. We tested and found the effect works for both male and female disclosers (See Supplemental Materials). Future research may explore other dimensions, such as the age and cultural background of the discloser. And although status (i.e., respect and admiration) and power (i.e., resource control) are distinct constructs, they often co-occur (e.g., Magee & Galinsky, 2008); that is, organizational leaders are often high in both status and power. Future work could therefore explore the "off-diagonals" (e.g., Blader, Shirako, & Chen, 2016; Fast, Halevy, & Galinsky, 2012) – i.e., what happens when a high-status but low-power individual makes sensitive self-disclosures relative to someone who is low in status but high in power. Relatedly, future work could also delve deeper into the types of sensitive self-disclosures that do versus do not confer an authenticity benefit for leaders. Here, we have operationalized "sensitive self-disclosures" as competence-related weaknesses and have shown authenticity gains to self-disclosing mild to moderate weaknesses. Future work could, for

example, examine what happens when leaders disclose other types of sensitive information, such as their moral transgressions.

Another future direction would be to examine how audience size influences the observed effect. Does self-disclosing weaknesses to a large audience (e.g., broadcasting) lead to higher or lower perceptions of authenticity than self-disclosing to a small audience (e.g., one person, or narrowcasting)? It may also help further rule out reciprocity as an alternative mechanism. Past research suggests that actors' motives differ across audience size: actors tend to focus on the audience when talking to one or two people (narrowcasting) but on the self when talking to a large group (broadcasting, Barasch & Berger, 2014). However, from recipients' side, whether recipients perceive an actor who discloses to a small group (vs. large group) of people as more authentic is an empirical question for future research.

Finally, conceptually, authenticity may have some overlap with the construct of warmth. In our studies, we demonstrated that, controlling for perceived warmth and competence, perceived authenticity predicted overall impressions and positive interpersonal outcomes, suggesting that perceived authenticity accounts for unique variance (Studies 2A and 4). However, future research is needed to tease apart the different aspects of warmth and authenticity. Such an investigation could further our understanding of the distinction between these two constructs and their differential effects on perception and behavior.

Conclusions

By making sensitive self-disclosures, leaders can enhance how authentic their followers perceive them to be, leading to positive interpersonal outcomes, and potentially organizational ones as well. Aside from the obvious costs of disclosing weaknesses, leaders may also reap surprising benefits from doing so.

References

Aronson, E., Willerman, B., & Floyd, J. (1966). The effect of a pratfall on increasing

interpersonal attractiveness. *Psychonomic Science*, *4*(6), 227–228.

Avolio, B. J., Gardner, W. L., Walumbwa, F. O., Luthans, F., & May, D. R. (2004). Unlocking

the mask: A look at the process by which authentic leaders impact follower attitudes and

behaviors. *The Leadership Quarterly*, *15*(6), 801–823.

Bannister, B. D. (1986). Performance outcome feedback and attributional feedback: Interactive

effects on recipient responses. *Journal of Applied Psychology*, *71*(2), 203–210.

Barasch, A., & Berger, J. (2014). Broadcasting and narrowcasting: How audience size affects

what people share. *Journal of Marketing Research*, *51*(3), 286–299.

Baugh, B. (1988). Authenticity revisited. *The Journal of Aesthetics and Art Criticism*, *46*(4),

477–487.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than

good. *Review of General Psychology*, *5*(4), 323–370.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and

Economic Behavior*, *10*(1), 122–142.

Blader, S. L., Shirako, A., & Chen, Y.-R. (2016). Looking out from the top: Differential effects

of status and power on perspective taking. *Personality and Social Psychology Bulletin*,

42(6), 723–737.

Blakely, G. L. (1993). The effects of performance rating discrepancies on supervisors and

subordinates. *Organizational Behavior and Human Decision Processes*, *54*(1), 57–80.

Brooks, A. W., Huang, K., Abi-Esber, N., Buell, R. W., Huang, L., & Hall, B. (2019). Mitigating

malicious envy: Why successful individuals should reveal their failures. *Journal of Experimental Psychology: General*, *148*(4), 667.

Bruk, A., Scholl, S. G., & Bless, H. (2018). Beautiful mess effect: Self-other differences in evaluation of showing vulnerability. *Journal of Personality and Social Psychology*, *115*(2), 192.

Carroll, G. R., & Wheaton, D. R. (2009). The organizational construction of authenticity: An examination of contemporary food and dining in the US. *Research in Organizational Behavior*, *29*, 255–282.

Cha, S. E., Hewlin, P. F., Roberts, L. M., Buckman, B. R., Leroy, H., Steckler, E. L., … Cooper, D. (2019). Being Your True Self at Work: Integrating the Fragmented Research on Authenticity in Organizations. *Academy of Management Annals*, *13*(2), 633–671.

Cheshin, A., Amit, A., & Van Kleef, G. A. (2018). The interpersonal effects of emotion intensity in customer service: Perceived appropriateness and authenticity of attendants' emotional displays shape customer trust and satisfaction. *Organizational Behavior and Human Decision Processes*, *144*, 97–111.

Chng, D. H. M., Kim, T.-Y., Gilbreath, B., & Andersson, L. (2018). Why People Believe in Their Leaders--or Not. *MIT Sloan Management Review*, *60*(1), 65–70.

Chng, D. H. M., Rodgers, M. S., Shih, E., & Song, X.-B. (2015). Leaders' impression management during organizational decline: The roles of publicity, image concerns, and incentive compensation. *The Leadership Quarterly*, *26*(2), 270–285.

Cianci, A. M., Hannah, S. T., Roberts, R. P., & Tsakumis, G. T. (2014). The effects of authentic leadership on followers' ethical decision-making in the face of temptation: An experimental study. *The Leadership Quarterly*, *25*(3), 581–594.

Clapp-Smith, R., Vogelgesang, G. R., & Avey, J. B. (2009). Authentic leadership and positive

psychological capital: The mediating role of trust at the group level of analysis. *Journal of*

Leadership & Organizational Studies, *15*(3), 227–240.

Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: A meta-analytic review.

Psychological Bulletin, *116*(3), 457.

Correll, S. J., Ridgeway, C. L., Zuckerman, E. W., Jank, S., Jordan-Bloch, S., & Nakagawa, S.

(2017). It's the conventional thought that counts: How third-order inference produces status

advantage. *American Sociological Review*, *82*(2), 297–327.

Cozby, P. C. (1972). Self-disclosure, reciprocity and liking. *Sociometry*, 151–160.

Cronin, C. (2019). Three ways authentic leaders inspire and retain employees. Retrieved from

https://www.entrepreneur.com/article/327220

Dalto, C. A., Ajzen, I., & Kaplan, K. J. (1979). Self-disclosure and attraction: Effects of intimacy

and desirability on beliefs and attitudes. *Journal of Research in Personality*, *13*(2), 127–

138.

De Angelis, M., Bonezzi, A., Peluso, A. M., Rucker, D. D., & Costabile, M. (2012). On

braggarts and gossips: A self-enhancement account of word-of-mouth generation and

transmission. *Journal of Marketing Research*, *49*(4), 551–563.

Derlega, V. J., Metts, S., Petronio, S., & Margulis, S. T. (1993). *Self-disclosure*. Newbury Park,

CA: Sage.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep

bidirectional transformers for language understanding. *ArXiv Preprint ArXiv:1810.04805*.

Ein-Gar, D., Shiv, B., & Tormala, Z. L. (2011). When blemishing leads to blossoming: The

positive effect of negative information. *Journal of Consumer Research*, *38*(5), 846–859.

Fast, N. J., Halevy, N., & Galinsky, A. D. (2012). The destructive nature of power without status. *Journal of Experimental Social Psychology*, *48*(1), 391–394.

Fine, G. A. (2003). Crafting authenticity: The validation of identity in self-taught art. *Theory and Society*, *32*(2), 153–180.

George, B., Sims, P., McLean, A. N., & Mayer, D. (2007). Discovering your authentic leadership. *Harvard Business Review*, *85*(2), 129.

Gibson, K. R., Harari, D., & Marr, J. C. (2018). When sharing hurts: How and why self-disclosing weakness undermines the task-oriented relationships of higher status disclosers. *Organizational Behavior and Human Decision Processes*, *144*, 25–43.

Grandey, A. A., Fisk, G. M., Mattila, A. S., Jansen, K. J., & Sideman, L. A. (2005). Is "service with a smile" enough? Authenticity of positive displays during service encounters. *Organizational Behavior and Human Decision Processes*, *96*(1), 38–55.

Green, P., Gino, F., & Staats, B. R. (2017). Shopping for confirmation: How disconfirming feedback shapes social networks. *Harvard Business School NOM Unit Working Paper*, (18–028).

Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, *111*(3), 364–371.

Gromet, D. M., & Pronin, E. (2009). What were you worried about? Actors' concerns about revealing fears and insecurities relative to observers' reactions. *Self and Identity*, *8*(4), 342–364.

Hahl, O., & Zuckerman, E. W. (2014). The denigration of heroes? How the status attainment process shapes attributions of considerateness and authenticity. *American Journal of*

Sociology, *120*(2), 504–554.

Hahl, O., Zuckerman, E. W., & Kim, M. (2017). Why elites love authentic lowbrow culture: Overcoming high-status denigration with outsider art. *American Sociological Review*, *82*(4), 828–856.

Hannah, S. T., Avolio, B. J., & Walumbwa, F. O. (2011). Relationships between authentic leadership, moral courage, and ethical and pro-social behaviors. *Business Ethics Quarterly*, *21*(4), 555–578.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Publications.

Herr, P. M., Kardes, F. R., & Kim, J. (1991). Effects of word-of-mouth and product-attribute information on persuasion: An accessibility-diagnosticity perspective. *Journal of Consumer Research*, *17*(4), 454–462.

Hirst, G., Walumbwa, F., Aryee, S., Butarbutar, I., & Chen, C. J. H. (2016). A multi-level investigation of authentic leadership as an antecedent of helping behavior. *Journal of Business Ethics*, *139*(3), 485–499.

John, L. K., Barasz, K., & Norton, M. I. (2016). Hiding personal information reveals the worst. *Proceedings of the National Academy of Sciences*, *113*(4), 954–959.

John, L. K., Blunden, H., & Liu, H. (2019). Shooting the messenger. *Journal of Experimental Psychology: General*, *148*(4), 644.

Jourard, S. M. (1959). Self-disclosure and other-cathexis. *The Journal of Abnormal and Social Psychology*, *59*(3), 428.

Kelly, A. E., & McKillop, K. J. (1996). Consequences of revealing personal secrets. *Psychological Bulletin*, *120*(3), 450.

Kingsley Westerman, C. Y., Reno, K. M., & Heuett, K. B. (2018). Delivering feedback:

Supervisors' source credibility and communication competence. *International Journal of Business Communication*, *55*(4), 526–546.

Laudert, E. (2018). Generation Z: Utilizing adaptive and authentic leadership to promote intergenerational collaboration at work.

Laurenceau, J.-P., Barrett, L. F., & Pietromonaco, P. R. (1998). Intimacy as an interpersonal process: The importance of self-disclosure, partner disclosure, and perceived partner responsiveness in interpersonal exchanges. *Journal of Personality and Social Psychology*, *74*(5), 1238.

Lehman, D. W., O'Connor, K., Kovács, B., & Newman, G. E. (2019). Authenticity. *Academy of Management Annals*, *13*(1), 1–42.

Leroy, H., Palanski, M. E., & Simons, T. (2012). Authentic leadership and behavioral integrity as drivers of follower commitment and performance. *Journal of Business Ethics*, *107*(3), 255–264.

Levashina, J., & Campion, M. A. (2007). Measuring faking in the employment interview: Development and validation of an interview faking behavior scale. *Journal of Applied Psychology*, *92*(6), 1638.

Linville, P. W. (1985). Self-complexity and affective extremity: Don't put all of your eggs in one cognitive basket. *Social Cognition*, *3*(1), 94–120.

Lynn, F. B., Podolny, J. M., & Tao, L. (2009). A sociological (de)construction of the relationship between status and quality. *American Journal of Sociology*, *115*(3), 755–804.

Lyubovnikova, J., Legood, A., Turner, N., & Mamakouka, A. (2017). How authentic leadership influences team performance: The mediating role of team reflexivity. *Journal of Business*

Ethics, *141*(1), 59–70.

Magee, J. C., & Galinsky, A. D. (2008). Eight social hierarchy: The self-reinforcing nature of

power and status. *Academy of Management Annals*, *2*(1), 351–398.

Moon, Y. (2000). Intimate exchanges: Using computers to elicit self-disclosure from consumers.

Journal of Consumer Research, *26*(4), 323–339.

Norman, S. M., Avolio, B. J., & Luthans, F. (2010). The impact of positivity and transparency on

trust in leaders and their perceived effectiveness. *The Leadership Quarterly*, *21*(3), 350–

364.

Rahimnia, F., & Sharifirad, M. S. (2015). Authentic leadership and employee well-being: The

mediating role of attachment insecurity. *Journal of Business Ethics*, *132*(2), 363–377.

Ramarajan, L. (2014). Past, present and future research on multiple identities: Toward an

intrapersonal network approach. *Academy of Management Annals*, *8*(1), 589–659.

Rego, A., Júnior, D. R., & e Cunha, M. P. (2015). Authentic leaders promoting store

performance: The mediating roles of virtuousness and potency. *Journal of Business Ethics*,

128(3), 617–634.

Rego, A., Vitória, A., Magalhães, A., Ribeiro, N., & e Cunha, M. P. (2013). Are authentic

leaders associated with more virtuous, committed and potent teams? *The Leadership

Quarterly*, *24*(1), 61–79.

Scopelliti, I., Loewenstein, G., & Vosgerau, J. (2015). You call it "self-exuberance"; I call it

"bragging": Miscalibrated predictions of emotional responses to self-promotion.

Psychological Science, *26*(6), 903–914.

Sergent, G. (2016). Who is the authenticity candidate of 2016? Yup, it's Donald Trump.

Retrieved from https://www.washingtonpost.com/blogs/plum-line/wp/2015/12/11/who-is-

the-authenticity-candidate-of-2016-yup-its-donald-trump/

Szalai, J. (2015). What Makes a Politician 'Authentic'? Retrieved from

https://www.nytimes.com/2016/07/10/magazine/what-makes-a-politician-authentic.html

Talbot-Zorn, J., & Marz, L. (2016). Donald Trump is not "authentic" just because he says the

bad things in his head. Retrieved from https://time.com/4519851/2016-election-authenticity/

Wang, D. S., & Hsieh, C. C. (2013). The effect of authentic leadership on employee trust and

employee engagement. *Social Behavior and Personality: An International Journal*, *41*(4),

613–624.

Wang, H., Sui, Y., Luthans, F., Wang, D., & Wu, Y. (2014). Impact of authentic leadership on

performance: Role of followers' positive psychological capital and relational processes.

Journal of Organizational Behavior, *35*(1), 5–21.

Worthy, M., Gary, A. L., & Kahn, G. M. (1969). Self-disclosure as an exchange process. *Journal

of Personality and Social Psychology*, *13*(1), 59.

Zhu, W., Avolio, B. J., Riggio, R. E., & Sosik, J. J. (2011). The effect of authentic

transformational leadership on follower and group ethics. *The Leadership Quarterly*, *22*(5),

801–817.

Zimmer, B. (2016). 'Authenticity' in the 2016 Campaign. Retrieved from

http://www.wsj.com/articles/authenticity-in-the-2016-campaign-1442591796

Zogby, J. (2016). Clinton, Trump and the battle for authenticity. Retrieved from

https://www.forbes.com/sites/johnzogby/2016/09/24/clinton-trump-and-the-battle-for-

authenticity/

Zukin, S. (2008). Consuming authenticity: From outposts of difference to means of exclusion.

Cultural Studies, *22*(5), 724–748.

Table 1

Examples of Disclosive and Non-Disclosive Posts from the Social Networking Platform

Disclosive	Non-Disclosive
I'm in my 40s. I find the "millennial job interview" videos on LinkedIn appalling. I am annoyed by the people who post them and furious when colleagues acknowledge them.	PwC seems to be pursuing a more metrics driven management strategy.... if I remember correctly Deloitte did something similar in years past. Lessons learned big D?
Am I the only F single ???? hoping to have a cute single fish mate. I swear all of my mates are either married or women :(	What is your best Deloitte University story?
D2 and feeling burned out. Lots of reflection about not leaving after manager promotion and not sure if I want the trade off of $ and life/family ahead - suggestions on stay/exit?	Do we need to take ipad off from its case for airport security or it's fine putting it in a container with the case?

Table 2

Means, Standard Deviations for Study 4 (N = 754)

| Disclosure sensitivity	Measures					
	Authentic	Competent	Warmth	Impression	Status (Self-report)	Status (Ladder)
Mildly sensitive	5.57 (1.10)	5.51 (0.98)	5.45 (0.98)	3.03 (1.74)	5.70 (1.02)	7.09 (1.50)
Sensitive	5.89 (0.99)	5.66 (0.95)	5.67 (0.99)	3.45 (1.52)	5.92 (0.97)	7.28 (1.44)
Moderately sensitive	5.66 (1.08)	5.42 (1.14)	5.34 (1.14)	2.75 (1.91)	5.69 (1.08)	7.01 (1.4)
Extremely sensitive	5.88 (0.99)	5.47 (0.93)	5.23 (0.93)	2.78 (1.76)	5.69 (0.96)	6.84 (1.56)
Exceptionally sensitive	5.94 (0.94)	5.25 (1.14)	5.25 (1.01)	2.79 (1.78)	5.71 (0.96)	6.89 (1.60)
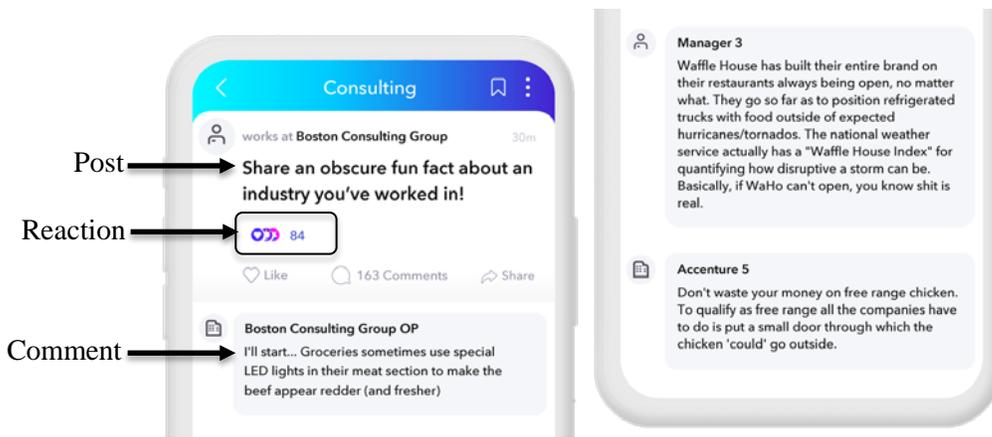
Figure 1. A Sample Screenshot of the Social Networking Platform from Study 1.
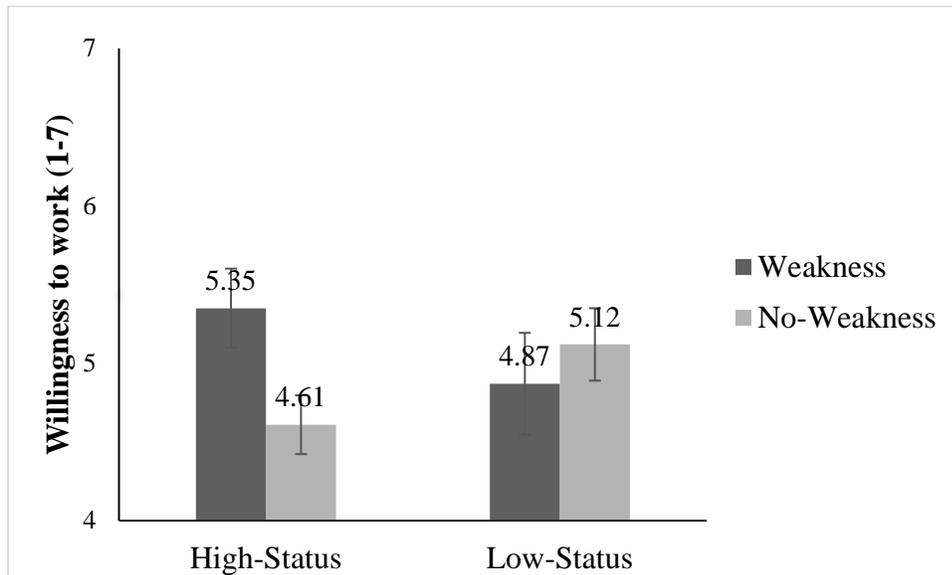
Figure 2. Willingness to work with the discloser by condition in Study 5.
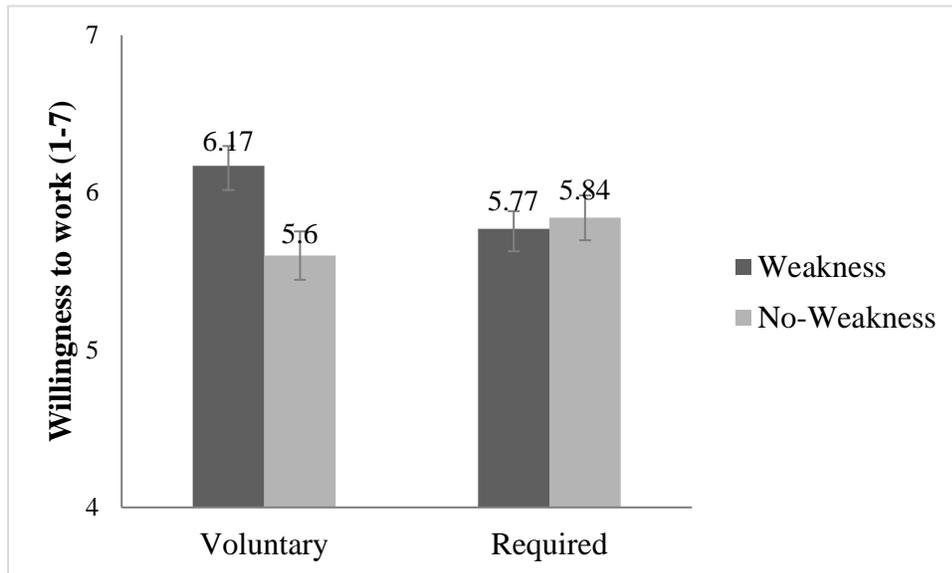
Notes. Error bars represent ±1 SE of the estimate.

Figure 3. Willingness to work with the discloser by condition in Study 6.

Notes. Error bars represent ±1 SE of the estimate.

ONLINE SUPPLEMENTAL MATERIAL

Supplemental Study I: Field Evidence

Results. As showed in the main text, self-disclosive posts and comments garnered significantly more positive reactions than non-disclosive ones. We coded a sample of our data and predicted labels for the remaining set using machine-learning methods. To determine whether performance of our machine-learning method affected the conclusion, we repeated our analysis using only the hand-coded data—that is, 3,484 posts and comments.

The results confirmed that self-disclosive posts and comments garnered more positive reactions ($M_{disclosive} = 7.68$, $M_{non-disclosive} = 5.89$, $t(1116.1) = -1.98$, $p < .05$), particularly likes ($M_{disclosive}=7.09$, $M_{non-disclosive} = 5.19$, $t(899.7) = -2.38$, $p < .05$). However, other types of reactions did not show any significant difference between the two groups.

We then applied the same analysis on the entire set of posts and comments separately. While there is no significant effect for posts ($M_{disclosive} = 9.52$, $M_{non-disclosive} = 8.71$, $t(1043.4) = -0.48$, $p = 0.63$), there is one for comments ($M_{disclosive} = 6.12$, $M_{non-disclosive} = 4.29$, $t(1137.3) = -4.62$, $p < .001$). Examining each reaction separately, except for "funny" ($M_{disclosive} = 0.23$, $M_{non-disclosive} = 0.25$, $t(1156.7) = 0.31$, $p = 0.76$), other reactions are used more frequently with self-disclosive comments (helpful: $M_{disclosive} = 0.10$, $M_{non-disclosive} = 0.04$, $t(1090.5) = -3.49$, $p < .001$; like: $M_{disclosive} = 5.6$, $M_{non-disclosive} = 3.94$, $t(1138.3) = -4.66$, $p < .001$; uplifting: $M_{disclosive} = 0.11$, $M_{non-disclosive} = 0.03$, $t(1081.1) = -3.47$, $p < .001$; smart: $M_{disclosive} = 0.07$, $M_{non-disclosive} = 0.03$, $t(1100.8) = -3.91$, $p < .001$).

Supplemental Study II

Method

   Procedure. Participants ($N = 300$ U.S. MTurk workers; 149 males; $M_{age} = 35.4$ years, $SD = 11.6$) were randomized to one of three conditions: a control condition, in which the CEO did not disclose a weakness, or one of two different experimental conditions, in which he disclosed a weakness. As in Study 2B, we created the experimental conditions by simply appending the disclosure of a weakness to the statement that all participants read (which was the same as in Study 2B).

   Specifically, in the *speaking weakness* condition, the CEO disclosed, as in Study 2B, that "Even if I am a manager of a multi-billion company, I am nervous about public speaking and I have a habit of cracking my knuckles." In the *teamwork weakness* condition, the CEO disclosed that "Even though I am a manager of a multi-billion company, I am nervous about teamwork and I have a habit of working alone." We do not expect a difference between the two weakness-disclosure conditions.

   Measures. We once again used the same measures of authenticity ($\alpha = .94$), competence ($\alpha = .93$), and overall impression, as in Study 2B.

Results

   Perceived authenticity. Perceptions of the CEO's authenticity were significantly different by condition ($F(2, 297) = 3.80$, $p = .023$). Specifically, relative to when he did not disclose weaknesses, the CEO was perceived as more authentic when he revealed that he was "nervous about public speaking" ($M_{speaking\_weakness} = 5.64$, $SD = 1.02$; $M_{control} = 5.22$, $SD = 1.18$), $t(192) = 2.55$, $p = .011$) or that he was "nervous about teamwork" ($M_{teamwork\_weakness} = 5.64$, $SD = 1.11$; $M_{control} = 5.22$, $SD = 1.18$), $t(196) = 2.05$, $p = .041$. Perceived authenticity was equivalent across

the two experimental conditions ($t(206)=.35$, $p = .73$).

Overall impressions. Overall impressions of the CEO were significantly different by condition ($F(2, 297) = 2.85$, $p = .059$). Specifically, relative to when he did not disclose weaknesses, overall impressions of the CEO were greater when he revealed that he was "nervous about public speaking" ($M_{speaking\_weakness}= 2.85$, $SD = 1.53$; $M_{control} = 2.24$, $SD = 2.04$), $t(192) = 2.39$, $p = .018$. Relative to when he did not disclose weaknesses, overall impressions of the CEO were directionally but not significantly greater when he revealed that he was "nervous about teamwork" ($M_{teamwork\_weakness} = 2.62$, $SD = 1.83$; $M_{control} = 2.24$, $SD = 2.04$), $t(196) = 1.40$, $p = .16$. Overall impressions were equivalent across the two experimental conditions ($t(206)=.99$, $p = .33$).

Perceived competence. Perceptions of the CEO's competence did not differ by condition ($F(2, 297) = 1.87$, $p = .16$).

Mediation. We conducted a mediation analysis with overall impressions as dependent variable to test mediation by perceptions of authenticity. A 10,000 sample bootstrap analysis (Model 4) showed that the index of mediation excluded zero (95% CI = [.015, .413]), suggesting a significant indirect effect (Hayes, 2017).

Supplemental Study III: Female Disclosers

We tested whether women leaders also benefit from self-disclosure of weaknesses, given the large body of research documenting differences between male and female leaders (for a review, see Eagly, 2005; Eagly & Karau, 2002; Rudman, 1998).

Method

Procedure. Participants (N = 399 U.S. MTurk workers, 232 males*; M_{age}* = 34.5 years, *SD* = 11.2) were randomly assigned to one of four conditions in a 2 (weakness vs. no-weakness) by 2 (gender of CEO: male, female) between-subjects design. As in Study 2B, we created the experimental conditions by simply appending the disclosure of a weakness ("I am nervous about public speaking and have a habit of cracking my knuckles") to the statement that all participants read.

Measures. We used the same measures as in Study 2B: authenticity ($α$ = .93), competence ($α$ = .85), overall impressions, and willingness to work with the leader.

Results

Perceived authenticity. A 2x2 ANOVA revealed only a significant main effect of disclosing a weakness ($F(1, 395) = 20.17$, $p < .001$), such that the manager was viewed as more authentic when he or she disclosed a weakness relative to when he or she did not do so ($M_{weakness\_disclosed}$ = 5.89, *SD* = .96; $M_{only\_strengths\_disclosed}$ = 5.41, *SD* = 1.15, $t(397) = 4.49$), $p < .0001$. There was no effect of gender ($F(1, 395) = 1.51$, $p = .22$) or two-way interaction ($F(1, 395) = .16$, $p = .69$).

Perceived competence. A 2x2 ANOVA on the perception of competence did not find any main effect or interaction.

Overall impressions. A 2x2 ANOVA revealed a significant main effect of disclosing a

weakness ($F(1, 395) = 7.00$, $p = .009$), such that the manager left a better overall impression when s/he disclosed a weakness relative to when s/he did not do so ($M_{weakness\_disclosed} = 3.15$, $SD = 1.67$; $M_{only\_strengths\_disclosed} = 2.68$, $SD = 1.82$), $t(397)=2.70$, $p = .007$. There was no effect of gender ($F(1, 395) = 3.50$, $p = .062$) or two-way interaction ($F(1, 395) = .005$, $p = .94$).

Willingness to work with the leader. A 2x2 ANOVA revealed only a significant effect of condition ($F(1, 395) = 3.77$, $p = .053$), such that disclosing a weakness increased willingness to work with the manager when s/he disclosed a weakness relative to when s/he did not do so ($M_{weakness\_disclosed} = 5.77$, $SD = 1.24$; $M_{only\_strengths\_disclosed} = 5.53$, $SD = 1.25$), $t(397) = 1.98$, $p = .048$. There was no effect of gender ($F(1, 395) = 2.65$, $p = .104$) or two-way interaction ($F(1, 395) = .59$, $p = .44$).

Mediation. We conducted a mediation analysis with overall impressions as the dependent variable to test mediation by authenticity while controlling for competence. A 10,000 sample bootstrap analysis (Model 4) showed that the index of mediation excluded zero (95% CI = [.363, .759]), suggesting a significant indirect effect. A mediation analysis by authenticity with willingness to work with the leader as the dependent variable while controlling for competence (Model 4) showed that the index of mediation excluded zero (95% CI = [.142, .411]), suggesting a significant indirect effect (Hayes, 2017).

Additional References

Eagly, A. H. (2005). Achieving relational authenticity in leadership: Does gender matter? *The Leadership Quarterly*, *16*(3), 459-474.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological review*, *109*(3), 573.

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

Supplemental Materials: The Correction Task in Study 3

Correction Task

On the next page, there will be a series of sentences that may or may not contain errors. If the sentence contains an error, you should select the best option to correct it. Some of the sentences do not have errors; if you do not believe the sentence has an error, choose the option that indicates that the sentence is correct as is. An example sentence and corresponding choices is below:

The cat are very fuzzy.

A) Change "fuzzy" to "fuzzier"
B) Change "are" to "is"
C) Remove the word "very"
D) The sentence is correct as is.

In this example, the correct answer is B.

You have 1 minute to get as many sentences correct as possible.

Q1. The man attempted to research the professor who he had met at the luncheon.

- ⚪ Change "who" to "whom"

- ⚪ Change "he" to "they"

- ⚪ Place a comma after "met"

- ⚪ The sentence is correct as is

Q2. Neither Ellen or Jane knew what to wear for their boss's holiday party.

- ⚪ Place a comma after "Jane"

- ⚪ Change "boss's" to "boss'"

- ⚪ Change "what" to "where"

- ⚪ Change "or" to "nor"

Q3. The judge looked completely uninterested as he stared out the window, ignoring the prosecutor's arguments.

- ⚪ Remove the comma from the sentence

- ⚪ Change "uninterested" to "disinterested"

- ⚪ Change "stared" to "staired"

- ○ The sentence is correct as is

Q4. My computer affected a much-needed transition from print sources to electronic documents.
- ○ Place a comma after "transition"
- ○ Place a semi-colon after "transition"
- ○ Change "transition" to "transitory"
- ○ Change "affected" to "effected"

Q5. There are less items in the grocery store next to our apartment than in the big store down the road.
- ○ Change "our" to "their"
- ○ Change "less" to "fewer"
- ○ Place a comma after "apartment"
- ○ Change "items" to "item"

Q6. Only one of these roses bloom in the summertime.
- ○ Change "roses" to "rose"
- ○ Change "bloom" to "blooms"
- ○ Place a comma after "roses"
- ○ The sentence is correct as is

Q7. The children have drank all of their milk in their cereal bowls.
- ○ Change "drank" to "drunk"
- ○ Change "drank" to drink"
- ○ Change "children" to "child"
- ○ Change "cereal" to "serial"

Q8. His mothers cabin is located in the woods in a small town in Vermont.
- ○ Change "mothers" to "mothers'"
- ○ Change "mothers" to "mother's"
- ○ Place a comma after "town"
- ○ Remove "the" from the sentence

Q9. Bones and diseases fascinate her so she has decided to become a doctor.

- ◯ Place a comma after "bones"

- ◯ Place a semi-colon after "her"

- ◯ Place a comma after "her"

- ◯ Place a comma after "diseases"

Q10. Jacob's favorite food is pizza, that also happens to be my favorite food.

- ◯ Remove the comma from the sentence

- ◯ Change "Jacob's" to "Jacobs"

- ◯ Change "food" to "foods"

- ◯ Change "that" to "which"