Fostering Perceptions of Authenticity via Sensitive Self-disclosure

Li Jiang

Leslie K. John

Reihane Boghrati

Maryam Kouchaki

Li Jiang, Department of Marketing, School of Business, George Washington University; Leslie John, Department of Negotiation, Organizations & Markets, Harvard Business School; Reihane Boghrati, Department of Information Systems, Arizona State University; Maryam Kouchaki, Department of Management & Organizations, Kellogg School of Management, Northwestern University. The research is self-funded, and does not have additional funding sources.

Correspondence concerning this article should be addressed to Li Jiang, Department of Marketing, School of Business, George Washington University, Washington DC, 20052. E-mail: lijiang1@gwu.edu.

Fort

Fostering Perceptions of Authenticity via Sensitive Self-Disclosure

Abstract

Leaders' perceived authenticity—the sense that leaders are acting in accordance with their "true self"—is associated with positive outcomes for both employees and organizations alike. How might leaders foster this impression? We show that sensitive self-disclosure, in the form of revealing weaknesses, makes leaders come across as authentic (Studies 1 & 2)—because observers infer that the discloser is not engaging in strategic self-presentation (Study 3). Further, the authenticity gains of sensitive self-disclosure have positive downstream consequences, such as enhancing employees' desire to work with the leader (Studies 4A and 4B). And, as our conceptual account predicts, these benefits emerge when the revealed weakness is made voluntarily (as opposed to by requirement) (Study 5), and are more pronounced if the disclosure is made by a relatively high-status person (Study 6). We also present anecdotal field evidence (Study 7) consistent with the causal effects identified in Studies 1-6.

Keywords: authenticity, weaknesses, self-disclosure, leaders' interpersonal perception

Public Significance Statement

When a leader self-discloses a weakness, s/he can be perceived as authentic, leading to positive downstream consequences, such as enhancing employees' desire to work with the leader. This research suggests that leaders can consider sensitive self-disclosure as a tool to achieve positive outcomes—for themselves, their employees, and the firm alike.

Fostering Perceptions of Authenticity via Sensitive Self-Disclosure

Authenticity has become increasingly important (Sergent, 2016; Szalai, 2015; Talbot-Zorn & Marz, 2016; Zimmer, 2016; Zogby, 2016). Research in organizational behavior indicates that employees prefer leaders whom they perceive to be authentic (e.g., Clapp-Smith et al., 2009; D. S. Wang & Hsieh, 2013; H. Wang et al., 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 et al., 2007; Lehman et al., 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 et al., 2004; Norman et al., 2010), perform better (Hannah et al., 2011; Leroy et al., 2012; Lyubovnikova et al., 2017; Rego et al., 2013, 2015), work harder (Hirst et al., 2016), and make more ethical decisions (Cianci et al., 2014; Zhu et al., 2011).

Despite these benefits, research also suggests that leaders struggle to come across as authentic (Hahl et al., 2017; Hahl & Zuckerman, 2014). 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 help leaders to come across as authentic? We propose that leaders can foster perceptions of authenticity by engaging in sensitive self-disclosure, which we operationalize in this context as revealing work-related weaknesses. In the following sections, we review prior literature that forms the basis of our predictions, and provide an overview of our empirical work.

Conceptual Development

Avoidance of Sensitive Self-Disclosure

People tend to shy away from revealing sensitive personal information (Bruk et al., 2018; De Angelis et al., 2012; Gromet & Pronin, 2009; John et al., 2016; Leary & Allen, 2011; Paulhus & Reid, 1991; Turnley & Bolino, 2001)—i.e., self-relevant information that makes a person vulnerable to being judged negatively by others (Derlega et al., 1993; Kelly & McKillop, 1996; Laurenceau et al., 1998; Moon, 2000). Within our context of interest—leaders' disclosures within the workplace—we define "sensitive self-disclosure" as revealing work-related weaknesses. This is because disclosing job-related weaknesses, such as not being good at public speaking, not being good at time management, or lacking a vision, plausibly makes a leader vulnerable to being judged negatively by followers—it may threaten followers' perceptions of that leader's ability to lead effectively. Indeed, shying away from sensitive self-disclosure can be sensible. For one, people tend to overweigh negative information relative to positive information (Baumeister et al., 2001; Herr et al., 1991). Prior work also indicates that revealing weaknesses can diminish others' perceptions of the discloser's status (Gibson et al., 2018).

At the same time, people tend to be willing to reveal favorable personal information—this is especially true in the workplace, where self-presentation concerns—the fundamental motive to be seen positively by others (Baumeister, 1982; Leary & Kowalski, 1990; Tetlock, 2002)—loom large. Accordingly, people tend to manipulate their images in an effort to be perceived in a desirable light (De Angelis et al., 2012; Leary & Allen, 2011; Paulhus & Reid, 1991; Turnley & Bolino, 2001); and doing so can result in social and material rewards (Gilmore & Ferris, 1989; Leary, 1996; Schlenker, 1975). For instance, in job interviews, people engage in extensive image creation, to the point of making up fictional stories to showcase their strengths

(Levashina & Campion, 2007). Relatedly, as organizational theorists have long noted, there is often a gap between the *frontstage*—i.e., a person's public persona—and their *backstage,* whereby the frontstage is manipulated by the actor to gain extrinsic rewards (Trilling, 1972; Turner, 1976).

Sensitive Self-Disclosure and Authenticity

We posit that the reticence to reveal unfavorable personal information comes with a cost: perceptions of inauthenticity. Self-presentational acts are often subject to assessments of authenticity (Buss & Briggs, 1984; Leary, 1993; Schlenker, 1975; Tesser & Moore, 1986), as observers infer whether the actor appears to be presenting her true self (Deci & Ryan, 2000, 2002; Kernis & Goldman, 2006; Lenton et al., 2013; Sedikides et al., 2017). If we only reveal our desirable qualities, we are only showing a very narrow "sample" of our true selves. Observers who notice such selective presentation may infer that the actor must be motivated to impress others and thus presenting an insincerely positive image to others.

In contrast, when a person engages in sensitive self-disclosure, observers may infer that the actor has not filtered out information. This may create the impression that the actor is revealing himself in a more complete, comprehensive, or unbiased way. As a result, we argue, observers perceive that actor as authentic. In making this proposition, we draw on seminal work in sociology on "staged authenticity"—the notion that access to "back regions" can enhance the intimacy, and perceived authenticity, of an experience (MacCannell, 1973)—as when, for example, a diner enters the kitchen area of a restaurant. Here, we posit that in interpersonal interactions, voluntarily allowing a person into one's "backstage" by revealing something sensitive, can foster perceptions of authenticity.

Central to our account, we propose that the capacity for sensitive self-disclosure to foster

perceptions of authenticity is driven by observers' inferences about the discloser's self-presentation motives. Prior work indicates that observers routinely make inferences about the motives that underlie others' behavior (Campbell & Kirmani, 2000; Heider, 1958; Pizarro et al., 2003). For example, when a salesperson flatters a consumer prior to making a sale, consumers perceive that salesperson as insincere—because consumers infer that the salesperson has an ulterior motive (i.e., the salesperson is being complimentary only to make the sale; Campbell & Kirmani, 2000). Conversely, when a person uses politically incorrect (vs. correct) language, she comes across as authentic because she is perceived to lack strategic motives (Rosenblum et al., 2020).

What inferences might observers make from a leader's sensitive self-disclosures? As in, what do observers perceive to be the leader's motive for engaging in sensitive self-disclosure? We posit observers to make inferences about the leader's self-presentation motives—or rather, the lack of such motives. Self-presentation is perceived as strategic behavior (Eastman, 1994), and acting strategically, we posit, is perceived as antithetical to behaving authentically. Thus, we propose that witnessing a leader self-disclose a weakness underscores an implicit assumption held by observers—that the person who discloses a weakness must *not* have filtered out information. Therefore, when a leader discloses a weakness, observers infer that that leader is not engaging in strategic self-presentation. As a result, the leader is perceived as relatively authentic.

Relevance to Prior Work

We build on prior work that points not to the pitfalls of sensitive self-disclosure—which may be what people typically focus on—but rather, *to its surprising benefits*; most notably, self-disclosure leads to liking (Collins & Miller, 1994; Cozby, 1972; Dalto et al., 1979; Jourard, 1959; Worthy et al., 1969). More recent work has elucidated that the capacity for self-disclosure

to foster liking is augmented by interpersonal attributions—that is, by inferences that the

discloser is engaging in sensitive self-disclosure in the interest of building rapport with the

recipient (Jiang et al., 2011; Kashian et al., 2017). Accordingly, this prior work has shown that in

dyadic conversations, the relationship between sensitive self-disclosure and liking is mediated by

such interpersonal attributions. Like this prior work, attributions are also central to our

theorizing. However, we posit the sensitive self-disclosure-authenticity link to be driven by

dispositional attributions—i.e., perceptions of the discloser's motivations for engaging in

sensitive self-disclosure.

The present research is also related to the Stereotype Content Model (SCM)—Fiske et

al.'s (2002) influential model of person perception, which posits two primary dimensions of

person perception: warmth and competence. From a SCM perspective, it is plausible that a

leader's sensitive self-disclosure could affect how warm and competent she comes across—and

these effects could potentially "crowd out" that of authenticity. Moreover, the SCM would seem

to treat the construct of "authenticity" as being a component of "warmth;" Fiske et al. (2002)

include "sincere" in their multi-item measure of warmth. However, more recent work suggests

that perceived authenticity is distinct from warmth. For example, Rosenblum, Schroeder, and

Gino (2020) showed that speaking in politically incorrect tones makes people come across as

authentic but *not* warm. Thus, we predict that sensitive self-disclosure will increase perceived

authenticity even when controlling for the SCM's two dimensions of person perception: warmth

and competence.

Finally, past research has shown that negative information can have positive effects in the

context of interpersonal attractiveness (Aronson et al., 1966; Collins & Miller, 1994), on the

management of malicious envy from peers (Brooks et al., 2019), and on enhancing the

effectiveness of persuasive appeals (e.g., two-sided messaging; see Crowley & Hoyer, 1994 as an example). We extend these findings and demonstrate that sensitive self-disclosure can enhance perceptions of authenticity. And, as we delineate in the next section, we further distinguish our account from related work by showing that it makes unique predictions about when sensitive self-disclosure will—versus will not—foster perceptions of authenticity.

Moderators

Voluntariness

We suggest that for leaders' disclosure of their weaknesses 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 and less blameworthy when they inadvertently, as opposed to intentionally, cause something bad to happen (Greene et al., 2009; Pizarro et al., 2003). And, of particular relevance to the domain of self-disclosure, 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 simply because they did not see the question) (John et al., 2016). Analogously, we propose that for leaders to reap the authenticity benefits of revealing weaknesses, 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' weaknesses; the act of *voluntary* self-disclosure is crucial to boosting perceptions of authenticity.

Status

We posit the authenticity gains from self-disclosing weaknesses in organizations to be pronounced for high-status individuals—i.e., leaders within the organization. First, this

prediction is rooted in work at the intersection of social identity and leadership. Specifically, as Giessner & van Knippenberg (2008) demonstrated, leaders are sometimes given a "license to fail" (Giessner & van Knippenberg, 2008): relative to low status individuals, high status individuals were treated more favorably after they failed to achieve a goal. Applied to the present context, this suggests that when leaders reveal weaknesses, they may be particularly poised to reap the benefits of doing so, and to avoid potential pitfalls. Note, too, that the focus is on *relative* status (for instance, in a consulting firm, a director is of high status compared to a junior associate, but is of relatively low status compared to a partner).

Second, prior work indicates that leaders are particularly driven to present themselves in a favorable light (Bolino et al., 2008; Peck & Hogue, 2018)—for example, when managers have information that makes them look bad, they are particularly likely to keep this information private (Harrison & Harrell, 1993). Thus, given leaders' particular reticence to reveal negative self-relevant information, we reason that when they do reveal such information, they are particularly likely to come across as authentic. Indeed, the results of a pilot study point to leaders' reticence to reveal weaknesses. We asked 110 full-time U.S. managers (M_{age} = 37.2 years, SD = 10.5; Male: 50.0%; White: 80.9%; Median income: $90,000 - $99,999) to write three pieces of self-relevant information: something favorable (i.e., something they are good at in the workplace), something neutral (i.e., hobbies), and something unfavorable (i.e., a workplace weakness). We then asked them which, if any, of these three facts about themselves they would include when introducing themselves to a new hire at work. Only 34.5% chose to disclose the weakness; by comparison, 96.3% chose to include the strength, and 64.5% chose to include the neutral fact (details in the supplement).

Overview of Studies

Our empirical package consists of four sections. In Section 1, we demonstrate the effect of a leader's disclosure of a weakness on perceptions of authenticity (Studies 1 and 2, Supplemental Studies 1A-1D) and the inferences that we posit to underlie it (Study 3). We begin by presenting vignette studies (Study 1 and S1A-S1D), followed by a more naturalistic study in which participants watch a video of a Google executive who either discloses or does not disclose a weakness within a self-introduction (Study 2). Pointing to the distinctiveness of the effect, we also show that it holds when controlling for liking (S1B), as well as perceived warmth and competence (Studies 1 and 2). Next, we show that the capacity for a leader's sensitive self-disclosure to foster authenticity is driven by the perception that he is _not_ engaging in strategic self-presentation (Study 3).

The goals of Section 2 are twofold: to further increase realism by using live interaction paradigms; and to assess positive behavioral consequences of leaders' self-disclosure of weaknesses—consequences that are downstream from the effect on perceived authenticity identified in Section 1. We show that followers are more likely to put their own earnings at risk in the hands of (Study 4A), and to choose to work with (Studies 4A & 4B), leaders who disclose weaknesses. Tying these patterns back to the basic effect identified in Section 1, in Section 2, we also show that these positive behavioral consequences are mediated by perceived authenticity.

Section 3 tests our theory-derived moderators: voluntariness (Study 5) and status (Study 6). Specifically, Study 5 shows that downstream positive consequences of revealing weaknesses in the workplace are limited to situations in which the disclosure is made voluntarily, as opposed to by requirement. Study 6 shows that these positive consequences are pronounced when the disclosure is made by a relatively high status person (Study 6). Again, we tie these patterns back

to the basic effect identified in Section 1; here, by measuring perceived authenticity and documenting moderated mediation.

In Section 4, we conclude by presenting anecdotal field evidence consistent with the causal effects identified in Sections 1-3. Specifically, using actual disclosures from a professional social networking app, we show that there is indeed a positive association between sensitive self-disclosure and reactions to those revelations (Study 7).

Transparency and Openness

In total—across the main manuscript and supplement—we report the results from twelve studies (N = 38,785; consisting of n = 3,712 from eleven experiments and n = 35,073 from a field study). Online subjects were recruited via Amazon Mechanical Turk or Prolific; Lab participants were recruited from a northeastern U.S. university. As for sample size, for our online experiments, we pre-set our sample size based on a power analysis that used the effect size from a preliminary study (Supplemental Study 1A; Cohen's $d = 0.50$) which indicated that 100 participants per condition would be required to have 95% power to detect an effect. Thus, in these studies (Studies 1-3, 4B, 5, and 6), we pre-set our target sample size to at least 100 per cell; specifically, we pre-set the sample size to 100 participants per condition in Studies 4B and 5, to 150 per condition in Studies 1, 3, and 6, and to 200 per condition in Study 2. For Study 4A, which was our in-person lab experiment (conducted prior to the power analysis), we pre-set our sample size to 50 participants per cell. For the field data (Study 7), we analyzed all of the data that our field partner gave us. We report all manipulations and measures; for brevity, some measures are only reported in the supplement. See Table 1 for a design overview of each study, and the supplement for more information about our supplemental studies, samples and designs. Most of our studies, not all of them, are pre-registered.

Section 1: Basic Effect and Mechanism

Study 1. Vignettes

In a series of vignette studies, we experimentally test the idea that when leaders disclose their weaknesses, they come across as authentic. Here, we present one of these studies in full. As for the others, we provide details on their relative contribution and a meta-analysis of their effect sizes; see the supplement for the full write-up of these additional studies.

Method

The study was a single-factor (weakness disclosed vs. control) between-subjects design; the primary outcome measure was perceived authenticity.

Participants. Full-time working professionals were recruited on Prolific ($N = 298$, 147 males; $M_{age} = 32.8$ years, $SD = 9.8$; White: 81.8%). We pre-registered our hypotheses, sample size, and measures (https://aspredicted.org/F9N_FWS).

Materials and Procedure. Participants imagined that they were a new employee of a (fictitious) company called RockInvest and they met different managers who they would have the opportunity to work with if they wanted to. Participants were randomly assigned to one of two conditions: a *control* condition, in which the manager did not disclose a weakness, or an *experimental* condition, in which he disclosed a work-related weakness.

Specifically, in the 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 experimental condition, we appended the following sentence, in which the manager disclosed a weakness: "Even if I am a manager of a multi-billion company, I am not good at public speaking. When I make a speech, my mouth gets dry and I sometimes start to panic."

Measures. We assessed perceived authenticity by asking participants to rate the CEO on six items ($\alpha = .95$: authentic, real, sincere, genuine, inauthentic (reverse-coded), phony (reverse-coded) on a 7-point scale from 1 (*not at all*) to 7 (*very much*). We combined these items to form a composite measure of perceived authenticity. The items were adapted from established perceived authenticity scales (see Cheshin et al., 2018; Gershon & Smith, 2020; Grandey et al., 2005; Hahl & Zuckerman, 2014). We also assessed competence ($\alpha = .92$: competent, efficient, intelligent) and warmth ($\alpha = .93$, warm, kind, easygoing) (Aaker et al., 2010; Fiske et al., 2007; Goodwin et al., 2014). The order of the authenticity, competence, and warmth measures was randomly assigned.

Results

Perceived authenticity. The manager was perceived as more authentic when he disclosed a weakness relative to when he did not ($M_{experimental} = 5.65$, $SD = 0.92$; $M_{control} = 4.97$, $SD = 1.11$), $t(296) = 5.69$, $p < .0001$, Cohen's $d = 0.56$.

Perceived competence and warmth. The manager was perceived to be just as competent when he disclosed a weakness relative to when he did not ($M_{experimental} = 5.41$, $SD = 0.85$; $M_{control} = 5.33$, $SD = 0.90$), $t(296) = .77$, $p = .44$, Cohen's $d = 0.09$; he was also perceived to be just as warm when he disclosed a weakness relative to when he did not ($M_{experimental} = 5.02$, $SD = 1.02$ vs. $M_{control} = 4.80$, $SD = 1.08$), $t(296) = 1.81$, $p = .07$, Cohen's $d = 0.25$. Moreover, the effect of condition on perceived authenticity held when controlling for both perceived competence and warmth ($t(294) = 6.31$, $p < .0001$), suggesting the effect of disclosure on

perceptions of authenticity is independent of perceptions of warmth and competence.

Finally, given that some of the prior work in person perception has treated authenticity as a sub-dimension of warmth, we also conducted a factor analysis of our authenticity and warmth measures. As further evidence of the distinctiveness of the authenticity construct, the factor analysis revealed two factors, with all of the authenticity items loading on one factor, and all of the warmth factors loading on the other factor (see the supplement for details).

Conceptual replications

Different weaknesses. Supplemental Studies 1A-1D are conceptual replications of the basic effect, showing that it emerges across a variety of weaknesses (see Table 1 for the weakness used in each study).

Sequence. In Supplemental Study 1A, we manipulated whether the weakness was presented at the beginning versus end of the leader's statement; the effect emerged in both cases. This robustness is noteworthy, as it distinguishes our effect from a related phenomenon, namely, how incorporating a small dose of negative information in product descriptions can lead to positive evaluations (Ein-Gar et al., 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, the present effect holds regardless of whether the weakness is disclosed upfront versus prefaced with the disclosure of neutral or desirable information.

Liking. Past research has demonstrated disclosure leads to liking (Collins & Miller, 1994). Thus, one important question is whether the authenticity pathway is independent of the liking pathway. Therefore, we measured liking in Supplemental Study 1B. A factor analysis indicated that the liking items and authenticity items each loaded onto their own distinct factor.

Further, the perceived authenticity pathway held even when controlling for liking (see the supplement for details).

Disclosure length. In Study 1, the leader's statement was longer in the experimental condition relative to the control condition. Therefore, we ran Supplemental Study 1C, in which we replicated the basic effect, this time keeping the length of the disclosure the same. Specifically, in Supplemental Study 1C, in the weakness-disclosed condition, the weakness *replaces* a (positive) piece of information in the control condition (as opposed to merely appending the weakness, as we did in Study 1). Study 6 is similar in this regard.

Gender. In Study 1, the leader was male. Given the large body of research documenting differences between male and female leaders (for a review, see Eagly, 2005; Eagly & Karau, 2002; Rudman, 1998), in Supplemental Study 1D, we tested whether the effect also emerges with female leaders; it did.

Meta-analysis. Finally, we performed a meta-analysis of the effect of self-disclosure of a weakness on perceived authenticity using the data from all of the above-mentioned studies. We used the R package meta (Schwarzer, 2007, v. 4.19-1) and used a random effects model by using the inverse variance method. The test of heterogeneity ($Q(6) = 2.85$, $p = .827$) was non-significant, suggesting that the studies consistently documented a significant condition effect of self-disclosure on perceived authenticity. The average effect size is Cohen's $d = 0.55$ (95% CI = [0.45, 0.65]). The results are in Figure 1.

Study 2. Increasing Realism

In Study 2 we invited a Google executive to record a video in which we instructed him to introduce himself, and to include a weakness. We did not give the executive guidance on what weakness to disclose; we wanted the stimuli to be as naturalistic as possible. We then edited the

video to create two clips; in the *experimental* condition we included the self-disclosed weakness,

and in the *control* condition we simply omitted it. We recruited working professionals, asked

them to imagine that they had recently joined the company, and randomly assigned them to view

one of the two versions of the video.

Method

The study was a single-factor (weakness disclosed vs. control) between-subjects design;

the primary outcome measure was perceived authenticity.

Participants. As outlined in our pre-registration (https://aspredicted.org/QJY_DNV), we

recruited full-time U.S. working professionals from Prolific ($N = 400$, 203 females; $M_{age} = 32.5$

years, $SD = 9.6$; White 76.4%).

Materials and Procedure. The executive was instructed to think about how he may

introduce himself to new employees at his company, and to include anything he would like to

this self-introduction. In addition, we asked him to disclose a weakness of himself; he disclosed

that he had joined the company "after applying to nearly 36 other roles and consequently

receiving 35 other rejections." See the supplement for the full video transcript. Between-subjects,

we manipulated whether this sentence was present. Next, we recruited working professionals and

instructed them to imagine that they had just joined a company and were meeting different

managers who they could choose to work with (or not). Participants were randomly assigned to

watch one of the two videos of the executive—the only difference being that in the *experimental*

condition, the manager disclosed a weakness; whereas in the *control* condition, he did not.

Measures. We used the same measures of authenticity ($\alpha = .94$), competence ($\alpha = .91$),

and warmth ($\alpha = .91$) as in Study 1; their order of administration was randomly assigned.

Results

Perceived authenticity. The manager was perceived as more authentic when he disclosed a weakness relative to when he did not ($M_{experimental} = 5.51$, $SD = 1.10$; $M_{control} = 5.24$, $SD = 1.16$), $t(398) = 2.41$, $p = .016$, Cohen's $d = 0.24$.

Perceived competence and warmth. The manager was perceived as just as competent when he disclosed a weakness relative to when he did not ($M_{experimental} = 6.06$, $SD = 0.83$; $M_{control} = 6.01$, $SD = 0.81$), $t(398) = .67$, $p = .50$, Cohen's $d = 0.06$. He was also perceived as just as warm when he disclosed a weakness relative to when he did not ($M_{experimental} = 4.85$, $SD = 1.16$ vs. $M_{control} = 4.68$, $SD = 1.22$), $t(398) = 1.43$, $p = .15$, Cohen's $d = 0.14$. Moreover, the effect of condition on perceived authenticity held when controlling for warmth and competence ($t(396) = 2.50$, $p = .013$).[1]

Study 3. Perceived Strategic Self-Presentation as Mechanism

We propose that when a leader discloses a weakness, it makes observers less likely to perceive that leader to be acting strategically, in turn fostering perceptions of authenticity. Thus, in Study 3, we test whether the effect of disclosing a weakness on perceived authenticity is mediated by inferences of strategic self-presentation. We use procedures similar to Study 1.

Method

The study was a single-factor (weakness disclosed vs. control) between-subjects design; the primary measures were perceived authenticity and our mediator, strategic self-presentation.

Participants. U.S. full-time working professionals from Prolific ($N = 300$, 146 males; $M_{age} = 31.9$ years, $SD = 8.6$; 78.5% White) participated; We pre-registered our hypotheses, sample size, and measures (https://aspredicted.org/blind.php?x=/TBG_TG8).

---

[1] We also measured perceived competence in some of the studies in Sections 2 and 3. Consistent with Studies 1 and 2, the results hold when we control for competence. Due to space constraints, however, we report these secondary results in the supplement.

Materials and Procedure. Participants imagined that they were a new employee of a (fictitious) company called RockInvest and were meeting different managers with whom they could choose to work. Participants were randomly assigned to one of two conditions: a *control* condition, in which the manager did not disclose a weakness, or the *experimental* condition, in which he disclosed a weakness. Specifically, in the control condition, participants were given the same description of the manager as in Study 1. In the experimental conditions, this sentence was appended: "Even though I have managed the company for many years, I struggle with adapting to new technologies, and as a manager I am not on top of technological changes." Participants then completed the dependent measures and provided demographic information.

Measures. We used the same measure of authenticity ($\alpha = .94$) as in Study 1, and added a measure of perceived strategic self-presentation, adapted from Rosenblum et al. (2020). Specifically, participants were asked: "to what extent does the manager's self-introduction seem to be strategic?" on a 7-point scale from 1 (*not at all*) to 7 (*very much*).

Results

Perceived authenticity. The manager was perceived as more authentic when he disclosed a weakness relative to when he did not ($M_{\text{experimental}} = 5.59$, $SD = 0.92$; $M_{\text{control}} = 4.79$, $SD = 1.25$), $t(298) = 6.37$, $p < .0001$, Cohen's $d = 0.73$).

Perceived strategic self-presentation. The manager was perceived as less likely to be engaging in strategic self-presentation when he disclosed a weakness relative to when he did not ($M_{\text{experimental}} = 5.08$, $SD = 1.18$; $M_{\text{control}} = 5.73$, $SD = 1.23$), $t(298) = -4.67$, $p < .0001$, Cohen's $d = 0.54$.

Mediation. Bootstrapping analyses (with 10,000 resamples) showed that perceived strategic self-presentation mediated the relationship between weakness disclosure and perceived

authenticity: the index of indirect effect excluded zero ($b = .100$, $SE = .046$, 95% CI = [.027, .211]), suggesting a significant indirect effect (Hayes, 2017). Specifically, self-disclosure of a weakness decreased perceptions of strategic self-presentation ($b = -.650$, $SE = .139$), $t(298) = - 4.67$, $p < .0001$, which in turn heightened perceived authenticity ($b = -.153$, $SE = .052$), $t(298) = - 2.96$, $p = .003$. Perceptions of strategic self-presentation explain 14.5% of the variance in perceived authenticity.

In sum, and consistent with our theorizing, Study 3 suggests that when a leader reveals a weakness, observers are less likely to think he is engaging in strategic self-presentation, which, in turn, increases perceived authenticity.

Section 2: Live Interactions and Behavioral Outcomes

So far, via vignette and video paradigms, we have documented that leaders can come across as authentic when they reveal weaknesses, and that this effect is driven by dampened perceptions of strategic self-presentation. And, pointing to the distinctiveness of the effect, it holds when controlling for perceptions of warmth and competence. The goals of Section 2 were twofold: to further increase realism by using live interaction paradigms; and to assesses positive behavioral consequences of leaders' self-disclosure of weaknesses—consequences downstream from the effect on perceived authenticity identified in Section 1. Specifically, we show that leaders' sensitive self-disclosure produces credible positive outcomes: employees are willing to risk their own money at the hands of the leader (Study 4A), and to choose to work with that leader for a subsequent task (Study 4B). Moreover, tying these patterns back to the basic effect identified in Section 1, in Section 2, we also show that the positive behavioral consequences of a leader's sensitive self-disclosure are mediated by perceived authenticity.

Study 4A: In-person Interaction Study

In Study 4A, participants engaged in a face-to-face, simulated employment task. In this interaction, the participant randomly assigned to the role of manager was privately instructed to disclose—or to not disclose—a weakness to the other participant, who was randomly assigned to the role of prospective employee. In an incentive compatible task, we then assessed whether employees were willing to entrust the manager with their money. Secondarily, we also assessed whether managers instructed to disclose a weakness would accurately predict that doing so would lead to positive behavioral outcomes (relative to managers randomly assigned to *not* disclose a weakness).

Method

The study was a single-factor (weakness disclosed vs. control) between-subjects design; the primary measures were perceived authenticity, willingness to work with the manager, and an incentive-compatible behavioral outcome.

Participants. We recruited students and community members ($N = 218$, 99 males; $M_{age} = 22.3$ years, $SD = 4.3$; White: 40.8%; Part-time employed: 42%; Full-time employed: 4%) to come to a lab at a northeastern university. Participants received a $15 base payment plus study earnings, as described below.

Materials and Procedure. In a simulated hiring task, we randomly assigned half of participants to the role of manager and the other half to the role of prospective employee, and randomly grouped participants into manager-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 their team and would assign the employee a ten-item "word correction" task. The word-correcting task served as a cover story. 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. 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 manager-employee 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 employees. Next, the manager assigned the employee the task and, using a stopwatch, gave the employee one minute to complete it.

Survey measures. After the task, participants returned to their individual cubicles. Each prospective employee assessed their manager's authenticity ($\alpha = .88$) as in Study 1, and their desire to work with the manager: "Would you want to be paired with this manager again for a subsequent task?" measured on a 7-point scale 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: "Do you think the prospective employee would want to be paired with you as a manager again for a subsequent task?" on a scale from 1 (*want to work with a different manager*) to 7 (*want to work with this manager*).

Finally, both managers and prospective employees indicated whether they knew each other before the experiment; three pairs did, and therefore were excluded from the data analysis, leaving 212 participants. The results are substantively equivalent when these three dyads are included.

Incentive-compatible behavioral outcome measure. Participants engaged in a trust game (Berg et al., 1995), which served as an incentive compatible measure of cooperative behavior (Cesarini et al., 2008; Epley et al., 2006). Essentially, we were interested in whether employees' positive assessments of a leader who self-discloses a weakness might manifest in a willingness to entrust the manager with their money; the trust game measures the extent to which people are willing to put their own money at risk by entrusting their counterpart (in this case, the manager) with it.

We explained the game to participants, telling them that the employee would receive an initial endowment of $3 (in quarters), and would have to decide how much, if any, of this money

to transfer to the manager. Any amount transferred would be tripled. Next, the manager would

decide how much, if any, of this tripled amount s/he would like to send back to the employee.

Participants were encouraged to ask questions or re-read the 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 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, and could earn as much as $9 (plus the $15 base payment). 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 their manager disclosed a weakness relative to when s/he did not ($M_{experimental} =$

5.43, $SD = 1.10$; $M_{control} = 4.92$, $SD = 1.09$), $t(104) = 2.42$, $p = .017$, Cohen's $d = 0.49$.

Willingness to work with the manager. 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 ($M_{experimental} = 5.41$, $SD = 1.34$; $M_{control} = 4.78$, $SD = 1.38$), $t(104) = 2.37$, $p = .019$,

Cohen's $d = 0.46$.

Incentive compatible behavioral outcome. Prospective employees transferred more

money when their manager disclosed a weakness relative to when s/he did not ($M_{experimental} =$

$2.39, $SD = 0.84$; $M_{control} = 2.00, $SD = 1.07$), $t(104) = 2.04$, $p = .043$, Cohen's $d = 0.41$.

Mediation. We conducted a mediation analysis with willingness to work as the dependent

variable to test mediation by authenticity. A 10,000-sample bootstrap analysis (Hayes, 2017,

Model 4) showed that the index of mediation excluded zero ($b = .330$, $SE = .146$, 95% CI = [.078, .652]), indicating a significant indirect effect. Perceived authenticity explained 30.2% of the variance in willingness to work. We observed similar mediation analyses results with incentive-compatible behavioral outcome—money allocation ($b = .130$, $SE = .069$, 95% CI = [.028, .315]) as dependent variable, again indicating a significant indirect effect. The direction of the mediations indicates that revealing a weakness increased perceived authenticity, in turn increasing both willingness to work with the manager, and money transferred. Perceived authenticity explained 11.8% of the variance in money allocation.

Managers. 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 ($M_{experimental} = 4.81$, $SD = 1.03$; $M_{control} = 4.56$, $SD = 1.19$), $t(104) = 1.16$, $p = .25$, Cohen's $d = 0.22$. This result suggests that would-be disclosers may be unaware of the benefits of sensitive self-disclosure: managers induced to disclose a weakness did not appear to anticipate that doing so would cause their employees to want to work with them. This result is consistent with the pilot study reported in the introduction, in which the majority of managers chose to *not* disclose a weakness in a self-introduction to a prospective employee.

Study 4B: Online Interaction Study with Working Professionals

Study 4B replicates and extends Study 4A in several ways. In Study 4B, we further enhanced realism in two ways. First, instead of giving participants a script, in Study 4B participants' weakness disclosures were self-generated. Second, we recruited working professionals (as opposed to students and community members, as we had in Study 4A); participants in Study 4B worked in various industries, in a wide range of professions. In addition, Study 4B examined the effect in the context of a manager-employee relationship rather than in

an interview context.

Method

The study was a single-factor (weakness disclosed vs. control) between-subjects design; the primary measures were employees' perceptions of the manager's authenticity, and their willingness to work with that manager.

Participants. We recruited working professionals from Prolific ($N = 400$, 187 males; $M_{age} = 30.5$ years, $SD = 10.7$; White: 72.1%); We pre-registered our hypotheses, sample size, and measures (https://aspredicted.org/Z8N_GND).

Materials and Procedure. We randomly assigned half of participants to the role of manager and the other half to the role of employee, randomly grouping them into manager-employee dyads in real-time. We used Qualtrics and SMARTRIQS software to perform this matching (c.f. Molnar, 2019). The matching procedure operates by having participants enter the survey at the same time—participants wait for up to two minutes until another participant joins; if no participant joins within two minutes, then the participant is thanked and paid for their participation. This procedure resulted in 80.3% of participants being matched (as indicated in our pre-registration, our target sample size was 400 *matched* participants). Importantly, this matching procedure occurred prior to randomization.

Each dyad was given four minutes to chat, ostensibly to get to know each other before engaging in a task together. Critically, it was during this "get-acquainted" chat session that we induced the managers to either disclose a weakness (*experimental* condition) or not (*control* condition). Thus, prior to the chat session, participants were given the following information.

Employees. At the start of the study, employees were told that they would be participating in a task together with the manager. Before the task, they would chat with their manager using a chat window.

Managers. Managers were instructed to start the chat with the employee by introducing themselves, and were randomly assigned to one of the two conditions. In both conditions, before entering the chat, managers were told: "First, we would like you to chat with the employee, to get to know each other a bit. As the manager, you will start the conversation by telling your employee a bit about yourself. Please chat in a natural way and make sure you respond to your employee." Further, managers were given specific information on what to include in their self-introduction, as follows:

"What to include in your introduction:

      1. Your profession (but not where you work).

      2. What you are good at.

For managers assigned to the *experimental* condition, there was a third bullet point, which read: "3. A work-related weakness." To help managers come up with a weakness, we further told these participants:

"For the work-related weakness, sometimes it's hard for people to come up with this. Here are some prompts that might help you come up with a weakness to reveal.

- Do you sometimes procrastinate? If so, you could say something like "I sometimes procrastinate and do things last minutes."

- Do you sometimes let your personal life interfere with your performance? If so, you could say something like "I have to admit that sometimes my personal life interferes with my job."

- Do you sometimes arrive late? If so, you could say something like "I am only

    human... occasionally I start work a little late."

Next, each pair was given four minutes to chat freely in a chat box in real time. After the

chat, employees completed several measures, described next. Managers provided demographic

information, were debriefed that there was no additional task, and paid.

Measures. Employees rated their managers' perceived authenticity, as in Study 4A.

Employees were also asked to choose whether they would like to work with the manager to

complete the task: "For a subsequent task you are going to complete, you have the option to

choose whether you want to work with this manager or to be paired with another manager. How

much are you willing to work with the manager in the subsequent task?" on a scale of 1 (*not at*

all) to 7 (*very much*). The order of authenticity and willingness to work questions were

counterbalanced. Next, all participants were asked: "Did you feel the chat to be natural?" (1 =

yes, 2 = *no*), provided demographic information, and were debriefed.

Results

Managers. All matched managers chatted with their employees. 92.3% of participants

indicated that they thought the chat to be natural. Compliance was high: in the control condition,

no managers disclosed a weakness, whereas in the experimental condition, 94.0% did so. A

research assistant coded the managers' disclosures; they disclosed a wide range of work-related

weaknesses: 22.3% disclosed weaknesses in time management (e.g., procrastinating, being late

for work), 19.1% in stress management, 18.1% in public speaking, 10.6% in social struggle,

7.4% in project management, 6.4% in the ability to focus, 5.3% in being patient, and 1.1% in

being overconfident. We adopt an intent-to-treat approach to data analysis, whereby all

employees were included in analysis, regardless of whether their manager complied. The results

remained significant when the non-compliers were removed from the analysis.[2]

Employees.

Perceived authenticity. Employees perceived their manager as more authentic when their manager was instructed to disclose a weakness relative to when s/he was not ($M_{experimental} = 4.96$, $SD = 1.35$; $M_{control} = 4.43$, $SD = 1.73$), $t(198) = 2.45$, $p = .015$, Cohen's $d = 0.34$.

Willingness to work with the manager for a subsequent task. Employees were more interested in working with their manager when their manager was instructed to disclose a weakness relative to when s/he was not ($M_{experimental} = 5.41$, $SD = 1.43$; $M_{control} = 4.90$, $SD = 1.91$), $t(198) = 2.14$, $p = .034$, Cohen's $d = 0.30$.

Mediation. We conducted a mediation analysis with willingness to work as the dependent variable to test mediation by authenticity. A 10,000-sample bootstrap analysis (Hayes, 2017; Model 4) showed that the index of mediation excluded zero ($b = .359$, $SE = .157$, 95% CI = [.083, .703]), suggesting a significant indirect effect. The direction of the mediation indicates that revealing a weakness increased perceived authenticity, which in turn increased willingness to work with the manager. Perceived authenticity explained 39.1% of the variance in willingness to work.

In sum, Studies 4A and 4B provide converging evidence that sensitive self-disclosure can make leaders come across as authentic, resulting in positive downstream consequences, such as a heightened interest in working for that leader.

Section 3: Moderators

---

[2] The results held when we excluded the 6.0% of employees whose manager did not comply with the disclosure manipulation: Perceived authenticity: $M_{experimental} = 4.98$, $SD = 1.35$; $M_{control} = 4.43$, $SD = 1.73$), $t(192) = 2.45$, $p = .015$, Cohen's $d = 0.35$. Willingness to work: $M_{experimental} = 5.45$, $SD = 1.42$; $M_{control} = 4.90$, $SD = 1.91$), $t(192) = 2.25$, $p = .025$, Cohen's $d = 0.33$. The mediation still held ($b = .370$, $SE = .160$, 95% CI = [.083, .706]).

In Section 3, we test theory-derived moderators: voluntariness (Study 5) and status (Study 6). First, we predicted that downstream positive consequences of revealing weaknesses in the workplace are limited to situations in which the disclosure is made voluntarily, as opposed to by requirement (Study 5). Second, we predicted that these consequences would be pronounced when the disclosure is made by someone of relatively high status (Study 6).

Study 5: Voluntariness

In Study 5, participants read a manager's disclosure and indicated their willingness to work with that manager, as well as their perceptions of that manager's authenticity. The study was a 2x2 between-subjects design; we manipulated the disclosure (weakness disclosed vs. no weakness disclosed), and the voluntariness of the disclosure (voluntary vs. required). We measured participants' perceptions of the manager's authenticity, and their willingness to work with the manager. We predicted an interaction such that interest in working for the manager would be heightened only in the condition in which the weakness was disclosed voluntarily, as opposed to by requirement. We also predicted that this effect would arise via gains in perceived authenticity—i.e., we predicted moderated mediation.

Method

Participants. U.S. MTurk workers ($N = 392$, 174 males*; $M_{age} = 29.8$ years, $SD = 12.3$; White: 80%; Median income: \$50,000-\$99,999) participated.

Materials and Procedure. Participants 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 like to climb mountains in Colorado and collect American folk art.

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

selected participants, appending the following self-disclosure: "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.) Specifically, when the manager had disclosed a

weakness, participants randomly assigned to the *required disclosure* condition were further told

that: "In the previous study, the individual was required to include some negative [positive] self-

relevant information in the introduction."

Measures. We measured participants' perceptions of the manager's authenticity ($\alpha = .92$)

as in Study 1. We measured participants' willingness to work with the manager. Specifically,

participants were asked "If you were looking for a job and were offered a job from this manager,

how likely would you accept the job and work for the manager?" on a 7-point scale from 1 (*not

at all*) to 7 (*very much*). The order of the two sets of questions was randomly assigned.

Results

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

($F(1, 388) = 17.43, p < .0001$): the manager was viewed as more authentic when he disclosed a

weakness relative to when he did not ($M_{weakness} = 5.57, SD = 1.04; M_{no\_weakness} = 5.11, SD = 1.06$),

$t(388) = 4.17, p < .0001$, Cohen's $d = 0.44$. However, this main effect was qualified by an

interaction ($F(1, 388) = 4.90, p = .027$) which suggested that 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 ($M_{\text{weakness\_volunteered}} = 5.75$, $SD = 0.96$; $M_{\text{no\_weakness\_volunteered}} = 5.06$, $SD = 1.06$), $t(193) = 4.52$, $p < .0001$, Cohen's $d = 0.68$. However, when managers were required to reveal a weakness, this difference disappeared ($M_{\text{weakness\_required}} = 5.39$, $SD = 1.17$; $M_{\text{no\_weakness\_required}} = 5.18$, $SD = 1.02$), $t(195) = 1.39$, $p = .17$, Cohen's $d = 0.18$.

Willingness to work with the leader. Mirroring the authenticity results, a 2x2 ANOVA revealed that disclosing a weakness only increased willingness to work with the manager when he did so voluntarily ($M_{\text{weakness\_volunteered}} = 5.85$, $SD = 1.04$; $M_{\text{no\_weakness\_volunteered}} = 5.51$, $SD = 1.11$), $t(193) = 2.19$, $p < .03$, Cohen's $d = 0.32$, as opposed to by requirement ($M_{\text{weakness\_required}} = 5.59$, $SD = 1.00$; $M_{\text{no\_weakness\_required}} = 5.66$, $SD = 1.15$), $t(195) = -0.43$, $p = .66$, Cohen's $d = 0.07$.

Moderated mediation. A moderated mediation analysis with willingness to work with the leader (10,000 sample bootstrap analysis, Hayes, 2017, Model 7) indicated that the index of moderated mediation excluded zero ($b = .301$, $SE = .132$, 95% CI = [.051, .570]), suggesting a significant indirect effect. Authenticity mediated the relationship between disclosure of weaknesses and willingness to work when disclosure was done voluntarily ($b = .434$, $SE = .098$, 95% CI = [.256, .641]) but not when disclosure was by requirement ($b = .133$, $SE = .101$, 95% CI = [-.059, .338]). Perceived authenticity explained 39.0% of the variance in willingness to work.

In sum, Study 5 suggests that for leaders to reap benefits from self-disclosing weaknesses, they must do so voluntarily.

Study 6: Status

In Study 6, we test the hypothesis that positive outcomes arising from self-disclosing

weaknesses mainly emerge for high-status disclosers. The study was a 2x2 between-subjects design; we manipulated the disclosure (weakness disclosed vs no weakness disclosed) and the discloser's status (*high* versus *low*). We measured participants' perceptions of the manager's authenticity and their willingness to work with that manager. We predicted an interaction, such that there would be heightened interest in working for a high-status, but not low-status, colleague who disclosed a weakness. And, as in Study 5, we also predicted that this effect would arise via gains in perceived authenticity—i.e., we predicted moderated mediation.

Method

      Participants. U.S. MTurk workers ($N = 614$, 310 males*; $M_{age} = 41.4$ years, $SD = 12.9$) participated.

      Materials and Procedure. Similar to prior studies, participants imagined that they were a new employee in the role of an analyst of a (fictitious) company called RockInvest and were meeting different people with whom they could choose to work. Participants were randomly assigned to one of four conditions. We manipulated discloser status by randomizing participants to imagine that Elis was either one of the senior managers (*high-status* condition) or one their peers (*low-status* condition). In all condition, Elis started with saying "I began my career 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." Participants were then randomly assigned to one of two conditions: a condition in which Elis disclosed a weakness (*weakness*), or a condition in which he did not disclose a weakness (*no weakness*). In the weakness condition, they read: "About myself, I need to fly several times a month, but I am afraid of flying." In the no weakness

condition, this sentence instead read: "About myself, I need to fly several times a month, and I enjoy flying."

Pretest. We conducted a pretest to ascertain whether the disclosed weakness was perceived as similarly sensitive across the status manipulation. We randomly assigned U.S. MTurk participants ($N = 391$, 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: "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*).

A 2x2 ANOVA revealed only a main effect of disclosing a weakness ($F(1, 387) = 18.31$, $p < .0001$): disclosing a fear of flying made Elis more vulnerable relative to when he did not do so ($M_{Weakness} = 2.15$, $SD = 0.85$; $M_{No\text{-}Weakness} = 1.79$, $SD = 0.81$), $t(387) = 4.38$, $p < .0001$, Cohen's $d = 0.44$. Importantly, there was no other main effect or interaction, suggesting that the weakness disclosure was seen as equally sensitive across status, hence, the disclosure manipulation was equally strong as a function of status.

Measures. Participants provided their perceptions of the discloser's authenticity ($\alpha = .96$) as in Study 1, and willingness to work with the manager as in Study 5 with one question "How much are you willing to work with Elis?".

Results

Perceived authenticity. A 2x2 ANOVA revealed a main effect of disclosing a weakness ($F(1, 610) = 36.68$, $p < .0001$): the person was viewed as more authentic when he disclosed a

weakness relative to when he did not ($M_{weakness} = 5.27$, $SD = 1.19$; $M_{no\_weakness} = 4.64$, $SD = 1.40$),

$t(612) = 6.05$, $p < .0001$. This main effect was qualified by an interaction ($F(1, 610) = 8.19$, p

$= .004$) which suggested that the authenticity benefit of revealing a weakness was stronger for

managers compared to non-managers. Specifically, disclosing a weakness boosted perceived

authenticity of both the high-status discloser ($M_{Weakness\_HighStatus} = 5.46$, $SD = 1.09$;

$M_{NoWeakness\_HighStatus} = 4.53$, $SD = 1.39$), $t(308) = 6.34$, $p < .0001$, Cohen's $d = 0.75$, and low-

status discloser ($M_{Weakness\_LowStatus} = 5.08$, $SD = 1.46$; $M_{NoWeakness\_LowStatus} = 4.75$, $SD = 1.34$),

$t(302) = 2.25$, $p = .025$, Cohen's $d = 0.24$. However, the effect size is greater for the high status

discloser.

Willingness to work with the discloser. A 2x2 ANOVA revealed a main effect of status

($F(1, 610) = 7.91$, $p = .005$), a main effect of weakness disclosure ($F(1, 610) = 15.83$, p

$< .0001$), qualified by an interaction ($F(1, 610) = 4.54$, $p = .034$): when the discloser was high

status, disclosing a weakness enhanced participants' interest in working with him

($M_{Weakness\_HighStatus} = 5.78$, $SD = 1.14$; $M_{NoWeakness\_HighStatus} = 5.09$, $SD = 1.50$), $t(308) = 4.34$, p

$< .0001$, Cohen's $d = 0.52$ – this effect was not observed when the discloser was low-status

($M_{Weakness\_LowStatus} = 5.22$, $SD = 1.47$; $M_{NoWeakness\_LowStatus} = 5.01$, $SD = 1.47$), $t(302) = 1.30$, p

$= .19$, Cohen's $d = 0.14$.

Moderated mediation. We conducted a 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 ($b = .460$, $SE = .175$, 95% CI $=$

[.129, .811]), suggesting a significant indirect effect (Hayes, 2017). Authenticity mediated the

relationship between disclosure of weaknesses and willingness to work for high status disclosers

(b = .736, SE = .123, 95% CI = [.510, .992]) and for low status disclosers (b = .277, SE = .128, 95% CI = [.026, .527]), but the effect is stronger for high status disclosers.

In sum, Study 6 suggests that positive outcomes arising from self-disclosing weaknesses are stronger for high-status disclosers.

Section 4: Field Evidence

In Section 4, we conclude by presenting anecdotal field evidence consistent with the causal effects identified in Sections 1-3. Specifically, data from a professional social networking platform reveal a positive association between sensitive self-disclosure and subordinates' reactions to those revelations. Another feature of this study is that we used a broader definition of sensitive self-disclosure: whereas the studies so far have operationalized sensitive self-disclosure in terms of revealing a weakness; here, we code any disclosure judged as making the discloser vulnerable to negative judgment, as sensitive (c.f. Derlega et al., 1993). As such, this study speaks to the generalizability of the effect. In addition, given the variation in status among users, we test for the moderating effect of status.

Study 7. Leaders' Sensitive Self-Disclosure on a Social Networking Platform

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. We assessed the positivity of reactions to senior leaders' posts as a function of whether the post contained a sensitive self-disclosure. This platform presents an ideal context for our research, because 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, as well as the reactions (i.e., "likes" and comments) 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 sensitive or

non-sensitive disclosure. Our primary interest was in testing whether reactions to, and comments

on, these posts and comments differed as a function of whether the content was self-disclosive.

We hypothesized that self-disclosive content would garner more positive reactions and

comments relative to less disclosive content. Second, although our dataset consisted only of

senior leaders' posts, there was nonetheless variation in status; Partners are the highest status in

this sample, followed by Directors, followed by Principals. Thus, this variation allowed for a

convergent test of our hypothesis, supported in Study 6, that the effect is moderated by status.

Specifically, as our dataset consists entirely of senior leaders' posts, our account holds that the

positive effect of self-disclosive posts may be observed across the ranks in the sample; however,

it also predicts an interaction, such that the positive effect increases with status.

Method

Data. Our dataset consists of posts and comments from senior leaders on a professional

social-networking platform, as well as reactions to, and comments on, those posts. A post is the

initiation of a new topic or thought. As on Facebook, fellow users can then respond to that post

in two ways: by reacting and/or commenting. A reaction entails pressing a button to choose one

of five possible reactions, all of which are positive: "like" (the default), "helpful," "funny,"

"uplifting," or "smart." A comment is a written response to the post (and, in turn, people can

react to, and comment on, comments). Unlike Facebook, this social network is exclusive; at the

time of data collection, it only granted access to those employed at certain firms within

consulting and advertising. The platform only requires that users reveal their rank and/or

employer. We requested the posts and comments from all senior leaders (operationalized as

Principals and above) within consulting firms, and the reactions and comments accompanying

these posts and comments. We focused on consulting firms because of its strict hierarchy

(advertising firms tend to have flatter organizational structures).

Our dataset consisted of 1,484 posts, which, collectively, garnered 159,221 reactions and

33,589 comments. 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 binary coding scheme to code all (1,484) posts and a random selection of 2,000

comments. We conducted a sensitivity analysis using GPower (Faul et al., 2009) to find the low

bound of effect size. The analysis, assuming two-tailed $\alpha = 0.05$ and 80% power, revealed a

minimum effect size of $d = 0.05$ (this sensitivity analysis is a rough estimate only, as our analysis

was non-parametric). Hereafter we refer to these posts and comments as "observations." This

quantity of observations was 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 sensitive self-disclosure. The team repeated this process twice for a total of approximately 30 randomly-selected observations. Next, 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 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 et al., 2018), the state-of-the-art deep-learning model in natural language processing which has recently received attention and been applied in different setting by researchers (Hartmann et al., 2021; Puranam et al., 2021). 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 training it on a randomly-selected sample of 80% of the 3,484 human-coded observations and applying it to the 20% holdout sample. 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 more (positive) reactions relative to non-self-disclosive ones ($M_{\text{disclosive}} = 6.98$, $SD = 16.15$; $M_{\text{non-disclosive}} = 4.44$, $SD = $

14.04), $t(1529.1) = -5.89$, $p < .001$, Cohen's $d = 0.17$). Looking at the different reactions separately; each reaction type, with the exception of "funny" was more prevalent for self-disclosive posts relative to no-self-disclosive ones: likes ($M_{disclosive} = 6.41$, $SD = 14.38$; $M_{non-disclosive} = 4.06$, $SD = 12.06$), $t(1517.1) = -5.94$, $p < .001$, Cohen's $d = 0.18$; helpful ($M_{disclosive} = 0.09$, $SD = 0.51$; $M_{non-disclosive} = 0.04$, $SD = 0.30$), $t(1477.9) = -3.41$, $p < .001$, Cohen's $d = 0.12$; uplifting ($M_{disclosive} = 0.10$, $SD = 0.68$; $M_{non-disclosive} = 0.03$, $SD = 0.29$), $t(1457.8) = -3.61$, $p < .001$, Cohen's $d = 0.13$; smart ($M_{disclosive} = 0.06$, $SD = 0.30$; $M_{non-disclosive} = 0.03$, $SD = 0.23$), $t(1494) = -3.4$, $p < .001$, Cohen's $d = 0.11$; funny ($M_{disclosive} = 0.32$, $SD = 1.90$; $M_{non-disclosive} = 0.27$, $SD = 2.87$), $t(1727.3) = -0.94$, $p = .35$, Cohen's $d = 0.02$). Results are equivalent when analyzing the two types of observations—posts versus comments—separately (see the supplement).

Further, there was a significant interaction between self-disclosiveness and status ($\beta = 2.32$, $p < .001$; Figure 2): self-disclosive posts and comments received more reactions when the discloser was of higher status. Specifically, self-disclosive posts by Partners were particularly likely to receive garner positive reactions ($M_{disclosive} = 12.26$, $SD = 23.29$; $M_{non-disclosive} = 6.65$, $SD = 16.91$, $t(256.61) = -3.74$, $p < .001$, Cohen's $d = 0.28$); followed by Directors ($M_{disclosive} = 6.17$, $SD = 15.36$; $M_{non-disclosive} = 3.89$, $SD = 14.34$, $t(912.85) = -4.22$, $p < .001$, Cohen's $d = 0.15$); followed by Principals ($M_{disclosive} = 5.18$, $SD = 9.89$; $M_{non-disclosive} = 4.19$, $SD = 11.47$, $t(374.54) = -1.81$, $p = .07$, Cohen's $d = 0.09$).

The above results are based on a binary coding scheme—each observation was coded as either containing, or not containing, a sensitive self-disclosure. In a supplementary analysis (see full details in the supplement), we asked two research assistants to code all 1,484 posts on a scale of 1 (*not sensitive at all*) to 5 (*very much sensitive*); their ratings were highly correlated ($r = .84$)

and disagreements were not more than one score apart. We then used the average of the two

coders' scores for each post to train a BERT model to predict the sensitivity scores for the

remaining data. There was a positive correlation between sensitivity score and positive reactions

($r = .08$, $p < .001$). Results held when analyzing the two types of observations—posts versus

comments—separately (see the supplement).

In sum, these results are consistent with our basic hypothesis, revealing an association

between leaders' propensity to engage in sensitive self-disclosure, and the positive reactions that

this activity appears to garner. That said, we acknowledge that positive reactions on social media

may not always be indicative of positive reactions in face-to-face interactions.

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 sensitive self-disclosure—and that this effect is

mediated by dampened perceptions of strategic self-presentation. Moreover, the increased

perceptions of authenticity arising from leaders' sensitive self-disclosures translate into broader

desirable outcomes (e.g., willingness to work for the leader). Also consistent with our conceptual

account, these positive effects are extinguished when the self-disclosure is involuntary and

pronounced when the self-discloser is of relatively high status.

Contribution to Theory

Most importantly, our research contributes to the leadership as well as authenticity

literatures. Indeed, Hewlin et al. (2020) called for more research focusing on the dynamics and

role of authenticity in the organizational context. The concept of authenticity (either experienced

and perceived) has been attracting a lot of attention from psychologists and management

scholars, as well as popular press (for recent reviews see Cha et al., 2019; Lehman et al., 2019; Sedikides et al., 2019). As we reviewed briefly, felt and perceived authenticity has been linked to numerous positive outcomes at work. However, questions related to what makes some people to be perceived as authentic, and does perceived authenticity matter or not are largely unanswered. Here, we examine the role of sensitive self-disclosure on perceived authenticity and document the positive consequences for perceived authenticity for high-status employees. Given that leaders are prone to being perceived as inauthentic (Hahl & Zuckerman, 2014) and given the organizational 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. Our theory suggests that for leaders to realize the benefits of sensitive self-disclosure, the disclosure has to be voluntary in nature. Importantly, we found that the positive effect of engaging in sensitive self-disclosure also emerge for female leaders.

We also contribute 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, and we demonstrate the effect of self-disclosure on perceptions of authenticity and subsequent outcomes over and above liking. Similar to Gibson et al. (2018), our work broadens the scope of self-disclosure from dyad relationships to organizationally-relevant settings. Complementing Gibson et al. (2018), our work suggests that perceived authenticity is a unique input to work-relevant interpersonal outcomes.

Accordingly, since we show a positive effect of self-disclosure, namely on perceptions of authenticity, that is distinct from the previously-documented effect on liking, we also demonstrate process evidence. Specifically, prior work has shown that the positive relationship

between self-disclosure and liking is driven by interpersonal attributions—that is, by inferences that the discloser is engaging in sensitive self-disclosure in the interest of having rapport with the recipient (Jiang et al., 2011; Kashian et al., 2017). In contrast, we show that the sensitive self-disclosure-authenticity link is driven by a different type of attribution, namely, dispositional ones: perceptions of the discloser's motivations for engaging in sensitive self-disclosure. Future work may further distinguish these two types of attribution processes.

Finally, our research contributes to the self-presentation literature by uncovering one way to soften the "braggart" image that is associated with self-promotion, and focusing on the work-related relationships in organizations. 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 et al., 2015). We suggest that by disclosing weaknesses, leaders may be able to come across as more authentic and generate more favorable outcomes without diminishing perceptions of their competence.

Contribution to Practice

Our research also offers practical implications. First and foremost, our research suggests that managers stand to benefit from revealing their weaknesses, at least to their subordinates. Interestingly, managers do not seem to intuit the benefits of such self-disclosure, raising the importance of disseminating our findings. Relatedly, disclosing weaknesses may not come naturally, especially to the high-achieving manager-type. Thus, managers may need to be quite intentional, at least at first, in applying our insights, pointing to a potential paradox: if the self-disclosure comes across as overly planned, would this impede the disclosure from coming across as authentic? Our work would suggest yes—to the extent that such "pre-meditated" self-disclosures come across as involuntary. Future work could explore the best ways for managers to

realize the benefits of sensitive self-disclosure, while mitigating its pitfalls.

Notably, we documented positive effects of sensitive self-disclosure even when controlling for liking. This finding may suggest that managers who are disliked could nonetheless benefit from sensitive self-disclosure, as it may at least make them come across as authentic. Finally, our research also speaks to the importance of self- and situational-awareness in realizing the benefits of sensitive self-disclosure. Specifically, Studies 6 and 7 imply that *relative* status matters, and this varies inherently from situation to situation: the higher relative status of the leader, it seems, the bigger the benefits of disclosing weaknesses. Thus it behooves leaders to be aware of their position in the status hierarchy when using this work prescriptively.

Future Research Directions

In addition to the future directions noted above, future work could identify additional moderators and boundary conditions. Although Study 7 operationalized "sensitive self-disclosure" as broader than disclosure of a workplace weakness (as in the other studies), additional research could systematically test other types of sensitive self-disclosures, to see our effects hold. For example, would they persist if the leader disclosed a disability or stigmatizing health condition? Relatedly, disclosure of serious weaknesses or immoral behavior may not forge positive impressions, for the authenticity gains of doing so may be offset by negative impressions. Relatedly, future work could assess whether knowing more diverse information about someone—i.e., that speaks to a person's multi-facetedness—may produce similar positive effects as we have documented here.

Future research could also assess whether self-disclosure of weaknesses works equally for different people. We found the effect to hold for both male and female disclosers (see Supplemental Study 1D) but future research could explore other dimensions, such as the age and

cultural background of the discloser. We showed the effect to emerge for high status disclosers

(Study 6); and the field data indicated that the higher status the discloser, the more pronounced

the effect. 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 et al., 2016; Fast et al., 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. In addition, even though we examined a wide range of sensitive self-

disclosures relevant to work, future work can explore self-disclosures of other vulnerabilities

such as disclosing of disabilities.

Another future direction would be to examine how audience size and audience structure

influence 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)? 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 the recipients' perspective, 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. Similarly, the audience structure might impact the observed effect. In Study 7

we focused on consulting firms because of their strict hierarchy, but the effects might weaken in

flatter organization structures.

Finally, conceptually, authenticity may have some overlap with the construct of warmth.

We demonstrated that our effect holds when controlling for perceived warmth and competence;

however, future research could aim to further distinguish warmth from authenticity. Such an

investigation could further our understanding of the distinction between these two constructs and

their differential effects on perception and behavior. Furthermore, the Stereotype Content Model

(SCM; Fiske et al., 2002) states that competence and warmth are two major dimensions of

person perception, but the two dimensions may not be the only dimensions. Our results suggest it

is possible that authenticity may be an additional dimension in addition to competence and

warmth. Future research may examine this possibility more closely using exploratory and

confirmatory factor analyses.

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. In sum, we conclude that despite their apprehension to reveal weaknesses, leaders

may reap surprising benefits from doing so.

References

Aaker, J., Vohs, K. D., & Mogilner, C. (2010). Nonprofits are seen as warm and for-profits as competent: Firm stereotypes matter. *Journal of Consumer Research*, *37*(2), 224–237.

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.

Barasch, A., & Berger, J. (2014). Broadcasting and narrowcasting: How audience size affects what people share. *Journal of Marketing Research*, *51*(3), 286–299.

Baumeister, R. F. (1982). Self-esteem, self-presentation, and future Interaction: A dilemma of reputation. *Journal of Personality*, *50*(1), 29–45. https://doi.org/https://doi.org/10.1111/j.1467-6494.1982.tb00743.x

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.

Bolino, M. C., Kacmar, K. M., Turnley, W. H., & Gilstrap, J. B. (2008). A multi-level review of impression management motives and behaviors. *Journal of Management*, *34*(6), 1080–1109.

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.

Buss, A. H., & Briggs, S. R. (1984). Drama and the self in social interaction. *Journal of Personality and Social Psychology*, *47*(6), 1310.

Campbell, M. C., & Kirmani, A. (2000). Consumers' Use of Persuasion Knowledge: The Effects of Accessibility and Cognitive Capacity on Perceptions of an Influence Agent. *Journal of Consumer Research*, *27*(1), 69–83.

Cesarini, D., Dawes, C. T., Fowler, J. H., Johannesson, M., Lichtenstein, P., & Wallace, B. (2008). Heritability of cooperative behavior in the trust game. *Proceedings of the National Academy of Sciences*, *105*(10), 3721–3726.

Cha, S. E., Hewlin, P. F., Roberts, L. M., Buckman, B. R., Leroy, H., Steckler, E. L., Ostermeier, K., & 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.

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.

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

Cronin, C. (2019). *Three ways authentic leaders inspire and retain employees*. Entrepreneur. https://www.entrepreneur.com/article/327220

Crowley, A. E., & Hoyer, W. D. (1994). An integrative framework for understanding two-sided persuasion. *Journal of Consumer Research*, *20*(4), 561–574.

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.

Deci, E. L., & Ryan, R. M. (2000). The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, *11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Deci, E. L., & Ryan, R. M. (2002). *Handbook of self-determination research*. University Rochester Press.

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*.

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.

Eastman, K. K. (1994). In the Eyes of the Beholder: An Attributional Approach to Ingratiation

    and Organizational Citizenship Behavior. *Academy of Management Journal*, *37*(5), 1379–

    1391.

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.

Epley, N., Caruso, E. M., & Bazerman, M. H. (2006). When perspective taking increases taking:

    reactive egoism in social interaction. *Journal of Personality and Social Psychology*, *91*(5),

    872.

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.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*

    Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*,

    41(4), 1149–1160.

Fine, G. A. (2003). Crafting authenticity: The validation of identity in self-taught art. *Theory and

    Society*, *32*(2), 153–180.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878.

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

Gershon, R., & Smith, R. K. (2020). Twice-told tales: Self-repetition decreases observer assessments of performer authenticity. *Journal of Personality and Social Psychology*, *118*(2), 307.

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.

Giessner, S. R., & van Knippenberg, D. (2008). "License to fail": Goal definition, leader group prototypicality, and perceptions of leadership effectiveness after leader failure. *Organizational Behavior and Human Decision Processes*, *105*(1), 14–35.

Gilmore, D. C., & Ferris, G. R. (1989). The effects of applicant impression management tactics on interviewer judgments. *Journal of Management*, *15*(4), 557–564.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*(1), 148.

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.

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.

Harrison, P. D., & Harrell, A. (1993). Impact of "adverse selection" on managers' project evaluation decisions. *Academy of Management Journal*, *36*(3), 635–643.

Hartmann, J., Heitmann, M., Schamp, C., & Netzer, O. (2021). The Power of Brand Selfies. *Journal of Marketing Research*, *58*(6), 1159–1177. https://doi.org/10.1177/00222437211037258

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

Heider, F. (1958). *The psychology of Interpersonal Relations*. Wiley.

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.

Hewlin, P. F., Karelaia, N., Kouchaki, M., & Sedikides, C. (2020). Authenticity at work: Its shapes, triggers, and consequences. In *Organizational Behavior and Human Decision Processes* (Vol. 158, pp. 80–82). Elsevier.

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.

Jiang, L. C., Bazarova, N. N., & Hancock, J. T. (2011). The disclosure--intimacy link in computer-mediated communication: An attributional extension of the hyperpersonal model. *Human Communication Research*, *37*(1), 58–77.

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.

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

Kashian, N., Jang, J., Shin, S. Y., Dai, Y., & Walther, J. B. (2017). Self-disclosure and liking in computer-mediated communication. *Computers in Human Behavior*, *71*, 275–283.

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

Kernis, M. H., & Goldman, B. M. (2006). *A Multicomponent Conceptualization of Authenticity: Theory and Research* (Vol. 38, pp. 283–357). Academic Press. https://doi.org/https://doi.org/10.1016/S0065-2601(06)38006-9

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.

Leary, M. R. (1993). The interplay of private self-processes and interpersonal factors in self-presentation. *Psychological Perspectives on the Self*, *4*, 127–155.

Leary, M. R. (1996). *Self-presentation: Impression management and interpersonal behavior*. Westview Publisher.

Leary, M. R., & Allen, A. B. (2011). Personality and persona: Personality processes in self-presentation. *Journal of Personality*, *79*(6), 1191–1218.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, *107*(1), 34.

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

Lenton, A. P., Bruder, M., Slabu, L., & Sedikides, C. (2013). How Does "Being Real" Feel? The Experience of State Authenticity. *Journal of Personality*, *81*(3), 276–289. https://doi.org/https://doi.org/10.1111/j.1467-6494.2012.00805.x

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.

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.

MacCannell, D. (1973). Staged authenticity: Arrangements of social space in tourist settings. *American Journal of Sociology*, *79*(3), 589–603.

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.

Molnar, A. (2019). SMARTRIQS: A simple method allowing real-time respondent interaction in Qualtrics surveys. *Journal of Behavioral and Experimental Finance*, *22*, 161–169.

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.

Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*, *60*(2), 307.

Peck, J. A., & Hogue, M. (2018). Acting with the best of intentions… or not: A typology and model of impression management in leadership. *The Leadership Quarterly*, *29*(1), 123–134.

Pizarro, D., Uhlmann, E., & Salovey, P. (2003). Asymmetry in Judgments of Moral Blame and Praise: The Role of Perceived Metadesires. *Psychological Science*, *14*(3), 267–272.

Puranam, D., Kadiyali, V., & Narayan, V. (2021). The Impact of Increase in Minimum Wages on

Consumer Perceptions of Service: A Transformer Model of Online Restaurant Reviews. *Forthcoming Marketing Science*.

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.

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.

Rosenblum, M., Schroeder, J., & Gino, F. (2020). Tell it like it is: When politically incorrect language promotes authenticity. *Journal of Personality and Social Psychology*, *119*(1), 75.

Rudman, L. A. (1998). Self-promotion as a risk factor for women: the costs and benefits of counterstereotypical impression management. *Journal of Personality and Social Psychology*, *74*(3), 629.

Schlenker, B. R. (1975). Self-presentation: managing the impression of consistency when reality interferes with self-enhancement. *Journal of Personality and Social Psychology*, *32*(6), 1030.

Schwarzer, G. (2007). meta: An R package for meta-analysis. *R News*, *7*(3), 40–45.

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.

Sedikides, C., Lenton, A. P., Slabu, L., & Thomaes, S. (2019). Sketching the contours of state

authenticity. *Review of General Psychology*, *23*(1), 73–88.

Sedikides, C., Slabu, L., Lenton, A., & Thomaes, S. (2017). State Authenticity. *Current Directions in Psychological Science*, *26*(6), 521–525. https://doi.org/10.1177/0963721417713296

Sergent, G. (2016). *Who is the authenticity candidate of 2016? Yup, it's Donald Trump*. Washington Post. 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'?* The New York Times. 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*. Time. https://time.com/4519851/2016-election-authenticity/

Tesser, A., & Moore, J. (1986). On the convergence of public and private aspects of self. In *Public self and private self* (pp. 99–116). Springer.

Tetlock, P. E. (2002). Social Functionalist Frameworks for Judgment and Choice: Intuitive Politicians, Theologians, and Prosecutors. *Psychological Review*, *109*(3), 451.

Trilling, L. (1972). *Sincerity and authenticity*. Harvard University Press.

Turner, R. H. (1976). The real self: From institution to impulse. *American Journal of Sociology*, *81*(5), 989–1016.

Turnley, W. H., & Bolino, M. C. (2001). Achieving desired images while avoiding undesired images: exploring the role of self-monitoring in impression management. *Journal of Applied Psychology*, *86*(2), 351.

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*. Wall Street Journal.

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

Zogby, J. (2016). *Clinton, Trump and the battle for authenticity*. Forbes.

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

Overview of Study Designs, Materials, and Key Measures

Study	Materials	Key dependent measures	Sample
Section 1: Effect of disclosure of a weakness on authenticity			
Study 1 (pre-registered)	Not good at public speaking: "I am not good at public speaking. When I make a speech, my mouth gets dry and I sometimes start to panic."	Authenticity	Prolific full-time employed
S1A	Speaking weakness: "I am nervous about public speaking and I have a habit of cracking my knuckles."	Authenticity	Mturk
S1B (three conditions)	Speaking weakness: same as S1A Technological weaknesses: "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."	Authenticity	Mturk
S1C (pre-registered) Keeping disclosure length constant	Technological weakness: "as a manager I struggle with keeping track of technological changes" Control: "as a manager I keep track of technological changes"	Authenticity	Prolific full-time employed
S1D	Speaking weakness: same as S1A Gender of the discloser: male leader vs. female leader	Authenticity	Mturk
Study 2 (pre-registered)	Real leader's video	Authenticity	Prolific full-time employed
Study 3 (pre-registered)	Technological weakness: same as S1B	Authenticity (strategic self-presentation as mediator)	Prolific full-time employed
Section 2: The outcomes of disclosure			
Study 4A	Dyad lab interaction study (with script): Speaking weakness as S1A	Authenticity Entrusting manager with money Willingness to work	Students and community members from a university lab
Study 4B (pre-registered)	Dyad online chat study (naturalistic setting): Self-generated weaknesses	Authenticity Willingness to work	Prolific full-time and part-time employed
Section 3: Moderators			
Study 5	Voluntary vs. requested disclosure: Speaking weakness as S1A	Authenticity Willingness to work	Mturk
Study 6	High vs. low status: Experimental: "About myself, I need to fly several times a month, but I am afraid of flying." Control: "About myself, I need to fly several times a month, and I enjoy flying."	Authenticity Willingness to work	Mturk
Section 4: Field Evidence			
Study 7	Field study	Likes and engagement of the posts	NA

Figure 1

Meta-analysis Results for Study 1, S1A-S1D

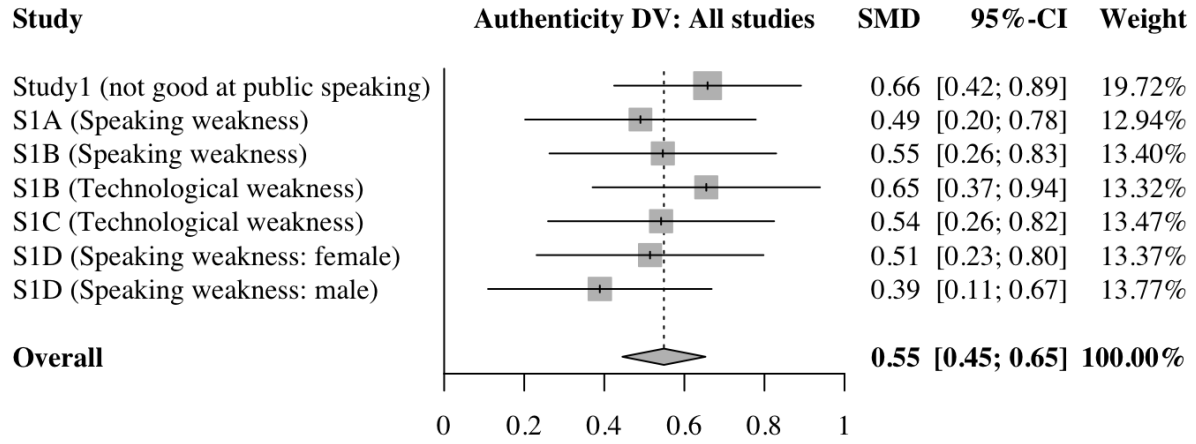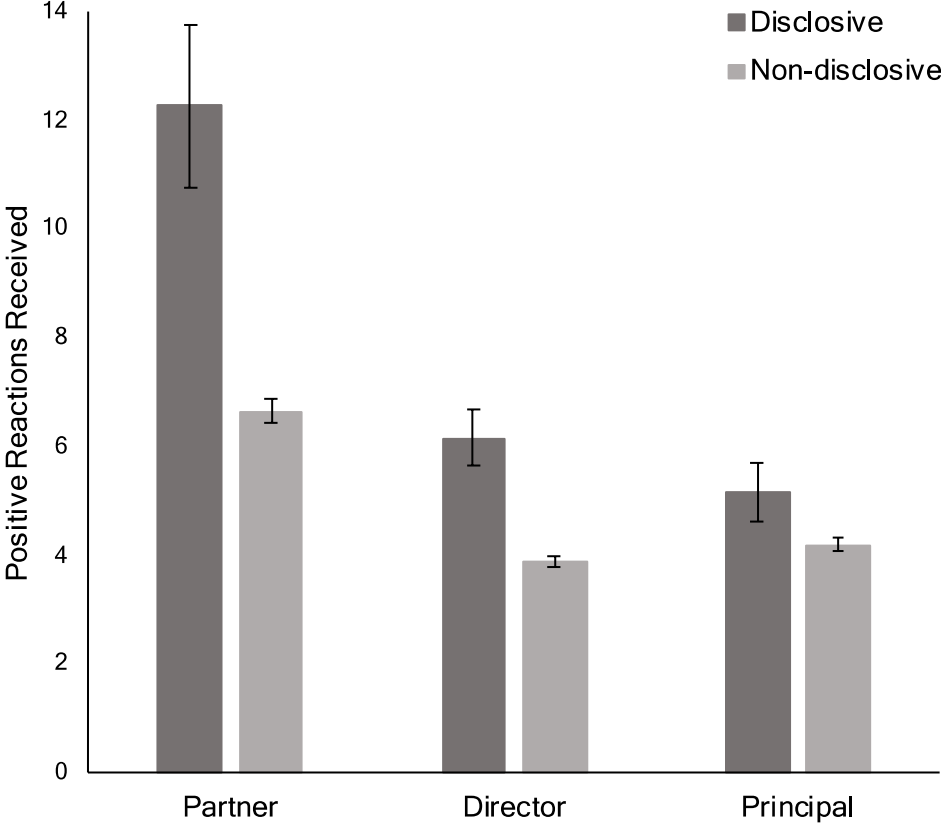| Study	Authenticity DV: All studies	SMD	95%-CI	Weight
Study1 (not good at public speaking)		0.66	[0.42; 0.89]	19.72%
S1A (Speaking weakness)		0.49	[0.20; 0.78]	12.94%
S1B (Speaking weakness)		0.55	[0.26; 0.83]	13.40%
S1B (Technological weakness)		0.65	[0.37; 0.94]	13.32%
S1C (Technological weakness)		0.54	[0.26; 0.82]	13.47%
S1D (Speaking weakness: female)		0.51	[0.23; 0.80]	13.37%
S1D (Speaking weakness: male)		0.39	[0.11; 0.67]	13.77%
Overall		**0.55**	**[0.45; 0.65]**	**100.00%**

Figure 2

Status Moderation Results for Study 7

Fostering Perceptions of Authenticity via Sensitive Self-Disclosure

ONLINE SUPPLEMENTAL MATERIAL

Supplemental Material I. Pilot: Managers' Intuition

Method

Participants

U.S. managers who manage at least four subordinates ($N = 110$, 55 males; $M_{age} = 37.2$ years, $SD = 10.5$; White: 80.9%; Median income: $90,000 - $99,999) were recruited from Prolific.

Materials and Procedure

Managers were asked to write three pieces of self-relevant information: something favorable (i.e., something they are good at in the workplace), something neutral (i.e., hobbies), and something unfavorable (i.e., a workplace weakness). For favorable information, participants were asked: "In the space below, please write down one of your skill-related advantages: e.g., what you are good at in the workplace." For neutral information, participants were asked: "In the space below, please write down one of your hobbies". For negative information, participants were asked: "We know no one is perfect. Everyone has weaknesses. In the space below, please write down one of your skill-related workplace weaknesses (i.e., a weakness you have in the workplace)." Then, participants were presented with the three pieces of information they have just written and were asked to choose which, if any, of these three facts about themselves they would include when introducing themselves to a new hire at work. Specifically, managers were told "Suppose you are meeting with a new hire and you want to introduce yourself and your group to him/her. How would you introduce yourself? Which information would you include

(choose all that apply)? Note: you can choose multiple answers." Participants next answered

demographic questions and the study concluded.

Results

All participants followed the instructions and wrote the three pieces of information. The

key dependent variable is the percentage of managers who included favorable, neutral and

unfavorable information in their introduction to the new hire. Only 34.5% chose to disclose the

weakness; by comparison, 96.3% chose to include the strength, and 64.5% chose to include the

neutral fact. These results suggest that managers do not intuit benefits of self-disclosing

weaknesses.

<center>**Supplemental Material II: Supplemental Studies 1A-1D**</center>

Pretest

We conducted a pretest to assess whether people deemed the weaknesses we intended to

use in Supplemental Studies 1A-1D to be sensitive self-disclosures. We presented pretest U.S.

MTurk participants ($N = 302$, 159 males; $M_{age} = 36.4$ years, $SD = 11.8$) with a quote from a CEO

of a (fictitious) company called RockInvest. Participants were randomly assigned to 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 judgment 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, Cohen's d = 0.34; ($M_{skills\_weakness}$ = 3.24, SD = 1.10; $M_{control}$ = 2.07, SD = 1.05), $t(193)$ = 7.51, p < .0001, Cohen's d = 1.09. 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). Supplemental Studies 1A and 1D test the former weakness; Supplemental Study 1C tests the latter weakness; Supplemental Study 1B tests both distinctly.

Supplemental Study 1A:

Speaking Weakness Presented at the Beginning of the Disclosure

Supplemental Study 1A was a single-factor two-condition (weakness disclosed vs. control) between-subjects design in which the CEO either disclosed (experimental condition) or did not disclose (control condition) a weakness; the primary outcome measure was perceived authenticity. Different from Study 1, we present the weakness at the beginning (rather than the end) of the disclosure to test for robustness. We predicted that participants would perceive the CEO as more authentic when he disclosed a weakness relative to when he did not. Similar to

Study 1, we also measured perceived competence and warmth.

Method

Participants

U.S. Amazon MTurk workers ($N = 191$, 102 males; $M_{age} = 34.7$ years, $SD = 11.2$) participated.

Materials and Procedure

Participants imagined that they were a prospective employee of a (fictitious) company called RockInvest and they met the CEO of company.

Participants were randomly assigned to one of two conditions: a *control* condition, in which the CEO did not disclose a weakness, or an *experimental* condition, in which he disclosed a weakness.

The control condition was the same as in the pretest.

In the experimental condition, the following sentence 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 disclosure, we used the same measure of perceived authenticity ($\alpha = .95$), **perceived competence ($\alpha = .92$), and perceived warmth ($\alpha = .89$) as in Study 1**. In this and all supplemental studies, the question order was randomized.

Results

Perceived authenticity

The CEO was perceived as more authentic when he disclosed a weakness relative to when he did not ($M_{experimental} = 5.72$, $SD = 1.09$; $M_{control} = 5.13$, $SD = 1.33$), $t(189) = 3.40$, $p = .0008$, Cohen's $d = 0.50$.

Perceived competence and warmth

The CEO was perceived as just as competent when he disclosed a weakness relative to when he did not ($M_{\text{experimental}}$ = 5.60, SD = 1.17; M_{control} = 5.58, SD = 1.15), $t(189)$ = 0.15, p = .88, Cohen's d = 0.02. The CEO was also perceived as just as warm when disclosed a weakness relative to when he did not ($M_{\text{experimental}}$ = 5.34, SD = 1.25 vs. M_{control} = 5.03, SD = 1.17), $t(189)$ = 1.68, p = .094, Cohen's d = 0.25.

Controlling for warmth and competence, there was still an effect of condition on perceived authenticity ($t(187)$ = 4.14, $p < .0001$), suggesting the effect of disclosure on perceptions of authenticity is independent of perceptions of warmth and competence.

Supplemental Study 1B:

Speaking Weakness and Technological Weakness

In Supplemental Study 1B, to test for convergent evidence of the effect, we added a second experimental condition, in which the CEO disclosed a work-related weakness: not being able to keep track of technological changes (*skills* weakness). The study was therefore a single-factor three-condition between-subjects design (*speaking* weakness disclosed vs. *skills* weakness disclosed vs. control); the primary outcome measure was perceived authenticity. We did not hypothesize a difference between the two weakness-disclosed conditions. To demonstrate the authenticity pathway is over and above liking, we also measured liking of the discloser in this study.

Method

Participants

We recruited U.S. Amazon MTurk workers (N = 300, 149 males; M_{age} = 35.4 years, SD = 11.6).

Materials and Procedure

Participants were randomly assigned 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 he disclosed a weakness.

Specifically, in the *control* condition, participants were told the same description of the CEO as in S1A. 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, 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**) and competence (α = **.91**), as in Study 1 and S1A. We used two items to measure liking of the discloser (r = **.77**). Specifically, participants indicated the extent to which they like, or they feel connected to the discloser, using a 7-point scale from 1 (*not at all*) to 7 (*very much*).

Results

Perceived authenticity

Perceptions of the CEO's authenticity were different by condition ($F(2, 297) = 13.6$, $p < .0001$). Specifically, relative to when he did not disclose weaknesses, the CEO was perceived as more authentic both in the *speaking weakness* condition ($M_{speaking\_weakness} = 5.74$, $SD = 1.04$; $M_{control} = 5.12$, $SD = 1.20$), $t(197) = 3.87$, $p < .0002$, Cohen's $d = 0.55$, and in the *skills weakness*

condition ($M_{skills\_weakness}$ = 5.88, SD = 1.12; $M_{control}$ = 5.12, SD = 1.20), $t(199)$ = 4.82, $p < .0001$,

Cohen's d = 0.65. Perceived authenticity did not differ between the *speaking weakness* and *skills*

weakness conditions ($t(198)$ = 0.93, p = .35, Cohen's d = 0.13).

Perceived competence

Perceptions of the CEO's competence did not differ by condition ($F(2, 297)$ = 0.71, p

= .49).

Liking

Liking of the CEO differed by condition ($F(2, 297)$ = 5.26, $p < .006$). Specifically,

relative to when he did not disclose weaknesses, liking of the CEO was higher in both the

speaking weakness condition ($M_{speaking\_weakness}$= 4.24, SD = 1.55; $M_{control}$ = 3.57, SD = 1.63),

$t(197)$ = 3.01, $p < .003$, Cohen's d = 0.42, and the *skills weakness* condition ($M_{skills\_weakness}$ = 4.13,

SD = 1.52; $M_{control}$ = 3.57, SD = 1.63), $t(199)$ = 2.53, p = .01, Cohen's d = 0.36. Liking did not

differ between the *speaking* and *skills weakness* conditions ($t(198)$ = .51, p = .61, Cohen's d =

0.07).

Controlling for competence and liking, there was still an effect of condition on perceived

authenticity for the speaking weakness ($t(196)$ = 3.17, $p < .002$) and technological weakness

($t(198)$ = 4.98, $p < .0001$), suggesting the effect of disclosure on perceptions of authenticity is

independent of perceptions of competence and liking.

Supplemental Study 1C:

Keep Track of Technology vs. Not Good at Keeping Track of Technology

Supplemental Study 1C was designed to control for the amount of information disclosed.

In this study, we kept the total amount of information constant across the conditions. In both the

control and the *experimental* conditions, the manager disclosed their technological skills. The

only difference was that in the control condition, the manager disclosed that he kept track of technological changes, whereas in the experimental condition, the manager disclosed that he was not able to keep track of technological changes. This study was a single-factor two-condition (weakness disclosed vs. control) between-subjects design; the primary outcome measure was perceived authenticity.

Method

Participants

Full-time working professionals from Prolific ($N = 200$, 101 males; $M_{age} = 35.3$ years, *SD* = 10.5; White: 86%) participated as outlined in the pre-registration (https://aspredicted.org/FXQ_PP9).

Materials and Procedure

Participants were randomly assigned to one of two conditions. We used the same scenario as S1B except we added a sentence at the end of the disclosure for both the *experimental* and the *control* condition. In the control condition, the CEO disclosed: "As a manager I keep track of technological changes." In the experimental condition, the CEO disclosed: "However, as a manager I struggle with keeping track of technological changes."

Measures

We once again used the same measures of authenticity ($\alpha = .95$) and competence ($\alpha = .89$) as in Study 1, S1A and S1B.

Results

Perceived authenticity

The CEO was perceived as more authentic when he disclosed a weakness relative to when he did not ($M_{experimental} = 5.61$, *SD* = 0.91; $M_{control} = 5.01$, *SD* = 1.15), $t(198) = 3.84$, *p*

$= .0002$, Cohen's $d = 0.58$.

Perceived competence

The CEO was perceived to be just as competent when he disclosed a weakness relative to when he did not ($M_{experimental} = 5.53$, $SD = 0.96$; $M_{control} = 5.61$, $SD = 1.05$), $t(198) = -0.61$, $p = .55$, Cohen's $d = 0.08$.

Controlling for competence, there was still an effect of condition on perceived authenticity ($t(197) = 5.58$, $p < .0001$) suggesting the effect of disclosure on perceptions of authenticity is independent of perceptions of competence.

Supplemental Study 1D: 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). This study was a 2 (weakness disclosed vs. control) x 2 (male leader vs. female leader) between-subjects design; the primary outcome measure was perceived authenticity. We did not have a pre-existing hypothesis on whether gender would moderate the effect.

Method

Participants

U.S. Amazon MTurk workers ($N = 399$, 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.

Materials and Procedure

As in Study 1, 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 1 and S1A-S1C: authenticity ($\alpha = $ **.93**) and

competence ($\alpha = $ **.85**).

Results

Perceived authenticity

A 2x2 ANOVA revealed only a 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 ($M_{experimental} = 5.89$, $SD = 0.96$; $M_{control} = 5.41$, $SD = $

1.15, $t(395) = 4.49$), $p < .0001$, Cohen's $d = 0.45$. There was no effect of gender ($F(1, 395) = $

1.51, $p = .22$) or two-way interaction ($F(1, 395) = 0.16$, $p = .69$).

Perceived competence

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

interaction.

Controlling for competence, there was still an effect of condition on perceived

authenticity ($t(394) = 3.96$, $p < .0001$), suggesting the effect of disclosure on perceptions of

authenticity is independent of perceptions of competence.

Supplemental Material III: Factor Analyses

For Study 1, we conducted a factor analysis of our authenticity and warmth measures.

We used the Varimax rotation method. We did not specify the number of factors. We found that

the authenticity items and the warmth items loaded on two distinct factors (Supplemental Table

1). These results suggest the authenticity construct is distinct from warmth.

Supplemental Table 1

Factor analysis of authenticity and warmth items used in Study 1

Category	Items	Factor 1	Factor 2
Authenticity Construct	Authentic	85*	40
	Real	80*	40
	Genuine	86*	38
	Sincere	78*	44
	Inauthentic	-86*	-22
	Phony	-85*	-20
Warmth Construct	Warm	34	83*
	Kind	36	83*
	Easygoing	22	83*

Note: Printed values are multiplied by 100 and rounded to the nearest integer. Values greater than 0.5 are flagged by an '*'.

Since liking is part of warmth and we measured liking directly in Supplemental Study

1B, we also conducted a factor analysis using authenticity items and liking items from

Supplemental Study 1B. We used the Varimax rotation method. Without specifying the number

of factors, we found that the authenticity items and the liking items loaded on two distinct factors

(Supplemental Table 2), suggesting the authenticity pathway is independent of liking.

Supplemental Table 2

Factor analysis of authenticity and warmth items used in Supplemental Study 1B

Category	Items	Factor 1	Factor 2
Authenticity Construct	Authentic	83*	41
	Real	82*	38
	Genuine	81*	38
	Sincere	74*	40
	Inauthentic	-84*	1
	Phony	-88*	-4
Liking	Like	23	88*

Category	Items	Factor 1	Factor 2
Construct	Connected	13	92*

Note: Printed values are multiplied by 100 and rounded to the nearest integer. Values greater than 0.5 are flagged by an '*'.

Supplemental Material IV: Video Transcripts in Study 2

Introduction: Control Condition

Hi, my name is Neil Hoyne. I'm the chief measurement strategist here at Google. The company, which you've probably heard of, was founded to help organize the world's information and make it universally accessible and useful. We have our search engine, YouTube, Gmail, google maps, and it seems like hundreds of other products that you probably have come across in your daily life.

Now I joined the company more than 10 years ago to help improve its display advertising campaigns. I eventually transitioned into the measurement ecosystem working with 16,000 of our top advertisers to understand the performance of their marketing spend with us. I've built products, machine learning models, and go to market strategies, but today spend most of my time simply helping advertisers make the most of their data.

Introduction: Weakness Condition

Hi, my name is Neil Hoyne. I'm the chief measurement strategist here at Google. The company, which you've probably heard of, was founded to help organize the world's information and make it universally accessible and useful. We have our search engine, YouTube, Gmail, google maps, and it seems like hundreds of other products that you probably have come across in your daily life.

Now, I joined the company more than 10 years ago to help improve its display advertising campaigns after applying to nearly 36 other roles and consequently receiving 35 other rejections. But eventually I transitioned into the measurement ecosystem working with 16,000 of our largest advertisers to understand the performance of their marketing spend with us. I've built products, machine learning models, and go to market strategies, but today spend most of my time simply helping advertisers make the most of their data.

Supplemental Material V

Supplemental Analysis for Study 7: Field Evidence

Binary Sensitivity Scale

As showed in the main text, self-disclosive posts and comments garnered 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 ($M_{disclosive} = 7.68$, $SD = 18.97$; $M_{non-disclosive} = 5.89$, $SD = 25.75$), $t(1116.1) = -1.98$, $p < .05$, Cohen's $d = 0.10$, particularly likes ($M_{disclosive} = 7.09$, $SD = 17.71$; $M_{non-disclosive} = 5.19$, $SD = 18.47$), $t(899.7) = -2.38$, $p < .05$, Cohen's $d = 0.11$. However, other types of reactions did not show any difference between the two groups.

We then applied the same analysis on the entire set of posts and comments separately. While there is no effect for posts ($M_{disclosive} = 9.52$, $SD = 23.22$; $M_{non-disclosive} = 8.71$, $SD = 38.83$), $t(1043.4) = -0.48$, $p = 0.63$, Cohen's $d = 0.02$, there is an effect for comments ($M_{disclosive} = 6.12$, $SD = 12.79$; $M_{non-disclosive} = 4.29$, $SD = 12.31$), $t(1137.3) = -4.62$, $p < .001$, Cohen's $d = 0.15$.

Examining each reaction separately, except for "funny" ($M_{disclosive} = 0.23$, $SD = 1.55$; $M_{non\text{-}disclosive}$ $= 0.25$, $SD = 1.69$), $t(1156.7) = 0.31$, $p = 0.76$, Cohen's $d = 0.01$, other reactions are used more frequently with self-disclosive comments (helpful: $M_{disclosive} = 0.10$, $SD = 0.58$, $M_{non\text{-}disclosive} = 0.04$, $SD = 0.30$, $t(1090.5) = -3.49$, $p < .001$, Cohen's $d = 0.13$; like: $M_{disclosive} = 5.6$, $SD = 11.52$, $M_{non\text{-}disclosive} = 3.94$, $SD = 11.16$, $t(1138.3) = -4.66$, $p < .001$, Cohen's $d = 0.15$; uplifting: $M_{disclosive} = 0.11$, $SD = 0.74$, $M_{non\text{-}disclosive} = 0.03$, $SD = 0.28$, $t(1081.1) = -3.47$, $p < .001$, Cohen's $d = 0.14$; smart: $M_{disclosive} = 0.07$, $SD = 0.35$, $M_{non\text{-}disclosive} = 0.03$, $SD = 0.22$, $t(1100.8) = -3.91$, $p < .001$, Cohen's $d = 0.14$).

Supplemental Analysis: Continuous Sensitivity

The primary analysis (reported in the main text) was based on a binary coding scheme—each observation was coded as either containing, or not containing, a sensitive self-disclosure. In the supplementary analysis, we asked two research assistants to code the 1,484 posts on a scale of 1 (*not sensitive at all*) to 5 (*very much sensitive*); their ratings were highly correlated ($r = .84$) and disagreements were not more than one score apart. We then used the average of the two coders' scores for each post to train a BERT model to predict the sensitivity scores for the remaining data. To test the model's predictive validity, we trained it on a randomly selected sample of 80% of the human-coded observations and tested it on the 20% holdout sample. Since the outcome variable is continuous, the performance is measured based on mean of absolute error (MAE). Our model reached MAE of 0.50—that is, on average, the predicted score and correct score were only 0.50 score apart.

Results of the continuous sensitive measure were consistent with those of the binary outcome. There was a positive correlation between sensitivity score and positive reactions ($r = .08$, $p < .001$). Specifically, the higher the sensitivity score, the more likely the given post or

comment was to receive likes reactions ($r = .09$, $p < .001$), helpful reactions ($r = .10$, $p < .001$), uplifting reactions ($r = .08$, $p < .001$), and smart reactions ($r = .07$, $p < .001$). There was no correlation for funny reactions ($r = .002$, $p = .66$). Analyzing posts and comments separately showed a similar pattern. More sensitive posts garnered more total positive reactions ($r = .09$, $p < .001$)—particularly likes ($r = .12$, $p < .001$). There is a marginally positive effect for uplifting and smart reactions (uplifting: $r = -.05$, $p = .05$; smart: $r = .04$, $p = .09$) and no effect for funny and helpful reactions (funny: $r = .02$, $p = .45$; helpful: $r = .02$, $p = .49$). Similarly, more sensitive comments garnered more total positive reactions ($r = .09$, $p < .001$). The effect held for all separate reactions, except funny reactions (helpful: $r = .11$, $p < .001$; like: $r = .09$, $p < .001$; uplifting: $r = .08$, $p < .001$; smart: $r = .07$, $p < .001$; funny: $r = .001$, $p = .89$). Further, there was a significant interaction between continuous self-disclosiveness score and status ($\beta = 0.92$, $p < .001$). We also measured the correlations within each status group. Results showed self-disclosive posts by Partners, Directors and Principals were all positively correlated with positive reactions, with the effect size greatest for Partners ($r = .131$, $p < .001$), followed by Directors ($r = .064$, $p < .001$), and by Principals ($r = .057$, $p < .001$).

Supplemental Material VI: Additional Measures

Here we report additional measures that we did not include in the manuscript.

Study 4A. Lab Interaction and Behavioral Outcomes

Method

In addition to the measures presented in the main manuscript, we also measured employees' perceived competence of the manager ($\alpha = $ **.88**) using the same measure as Study 1, and overall impressions of the manager on a scale from -5 (*very unfavorable*) to 5 (*very favorable*). 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*). As a cover story, we asked managers: "What is your overall impression of the prospective employee?" on a scale 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).

Results

Employees.

Perceived competence. Prospective employees did not perceive their manager as more competent when their manager disclosed a weakness relative to when s/he did not ($M_{experimental}$ = 5.18, SD = 1.02; $M_{control}$ = 5.10, SD = 1.17), $t(104)$ = 0.37, p = .71, Cohen's d = 0.07.

Overall impressions. Prospective employees showed directional but non-significant differences in overall impressions of their manager as a function of disclosure ($M_{experimental}$ = 3.00, SD = 1.50; $M_{control}$ = 2.43, SD = 2.00), $t(104)$ = 1.66, p = .10, Cohen's d = 0.32.

Mediation. We conducted a mediation analysis with overall impressions as the dependent variable to test mediation by authenticity. A 10,000-sample bootstrap analysis (Hayes, 2017, Model 4) showed that the index of mediation excluded zero (b = .520, SE = .237, 95% CI = [.115, 1.056]), suggesting a significant indirect effect. The direction of the mediation indicates that revealing a weakness led to greater perceptions of authenticity, which in turn contributed to greater overall impressions of the manager. Controlling for perceived competence, the effect still held (b = .273, SE = .125, 95% CI = [.075, .575]). Perceived authenticity explained 41.1% of the variance of overall impressions.

In the manuscript, we report mediation analyses with willingness to work and money

allocation as the dependent variables to test mediation by authenticity. Controlling for perceived competence, the index of indirect effect still excluded zero (willingness to work: $b = .229$, $SE = .109$, 95% CI = [.066, .515]; money allocation: $b = .158$, $SE = .080$, 95% CI = [.039, .364]), suggesting significant indirect effects.

Managers.

Overall impressions. 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 ($M_{experimental} = 2.10$, $SD = 1.33$; $M_{control} = 1.76$, $SD = 1.59$), $t(104) = 1.18$, $p = .24$, Cohen's $d = 0.23$. There was also no difference in managers' overall impressions of the paired employees ($M_{experimental} = 2.11$, $SD = 1.71$; $M_{control} = 2.02$, $SD = 1.86$), $t(104) = .26$, $p = .79$, Cohen's $d = 0.05$. We also did not expect this measure to differ, as the purpose of the measure is to fulfill the cover story.

Study 5: Voluntariness

Method

In addition to the measures presented in the main manuscript, we also measured participants' perceived competence of the manager ($\alpha = $ **.88**) using the same measure as in Study 1, and overall impressions of the manager as in Study 4A.

Results

Perceived competence. A 2x2 ANOVA revealed no main effects or interaction.

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 he disclosed a weakness relative to when he did

not do so ($M_{weakness\_volunteered} = 2.75$, $SD = 1.67$; $M_{no\_weakness\_volunteered} = 2.32$, $SD = 1.76$, $t(388) =$ 2.52, $p = .012$, Cohen's $d = 0.25$). 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 ($M_{weakness\_volunteered} = 2.98$, $SD = 1.60$; $M_{no\_weakness\_volunteered} = 2.14$, $SD = 1.73$), $t(193) = 3.44$, $p = .0007$, Cohen's $d = 0.50$, as opposed to by requirement ($M_{weakness\_required} = 2.53$, $SD = 1.70$; $M_{no\_weakness\_required} = 2.50$, $SD = 1.72$), $t(195) = .12$, $p = .90$, Cohen's $d = 0.02$.

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. A 10,000-sample bootstrap analysis (Hayes, 2017, Model 7) showed that the index of moderated mediation excluded zero ($b = .290$, $SE = .152$, 95% CI = [.004, .603]), suggesting a significant indirect effect. Authenticity mediated the relationship between disclosure of weaknesses and overall impressions when disclosure was done voluntarily, ($b = .441$, $SE = .124$, 95% CI = [.218, .709]) but not when disclosure was by requirement ($b = .152$, $SE = .114$, 95% CI = [-.058, .386]). Controlling for competence, the moderated mediation also held ($b = .572$, $SE = .247$, 95% CI = [.088, 1.067]): authenticity mediated the relationship between disclosure of weaknesses and overall impressions when disclosure was done voluntarily ($b = .825$, $SE = .182$, 95% CI = [.485, 1.202]) but not when disclosure was by requirement ($b = .253$, $SE = .188$, 95% CI = [-.112, .636]). Perceived authenticity explained 56.7% of the variance of overall impressions.

In the manuscript, we report mediation analysis with willingness to work as the dependent variable to simultaneously test moderation by voluntariness and mediation by perceptions of authenticity. Controlling for competence, the effect still held. Authenticity

mediated the relationship between disclosure of weaknesses and willingness to work when disclosure was done voluntarily ($b = .184$, $SE = .057$, 95% CI = [.087, .655]) but not when disclosure was by requirement ($b = .063$, $SE = .049$, 95% CI = [-.025, .311]).

Study 6: Status

Method

In addition to the measures presented in the main manuscript, we also measured participants' perceived competence of the manager ($\alpha = .94$) using the same measure as in Study 1.

Results

Perceived competence. A 2x2 ANOVA revealed only a main effect of status ($F(1, 610) = 6.07$, $p = .014$.

In the manuscript, we report mediation analysis with willingness to work as the dependent variable to simultaneously test moderation by status and mediation by perceptions of authenticity. Controlling for competence, the moderated mediation still held ($b = .267$, $SE = .108$, 95% CI = [.064, .489]).

Supplemental Study 1A.

Results

Overall impressions. Participants had a better impression of the CEO when he disclosed a weakness relative to when he did not ($M_{experimental} = 2.91$, $SD = 1.74$; $M_{control} = 2.30$, $SD = 2.03$), $t(189) = 2.22$, $p = .028$, Cohen's $d = 0.32$.

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 ($b = .725$, $SE = .234$, 95% CI = [.298, 1.215]). The direction

of the mediation indicates that revealing a weakness led to greater perception of authenticity, which in turn contributed to better overall impressions of the manager. This mediation held when controlling for perceptions of competence and warmth ($b = .208$, $SE = .070$, 95% CI = [.093, .373]), suggesting that authenticity explains unique variance in the relationship between leaders' sensitive self-disclosures and overall positive impressions. Perceived authenticity explained 61.9% of the variance of overall impressions.

Supplemental Study 1B. Speaking Weakness and Technological Weakness

Method

In addition to the main measures, participants provided their overall impressions of the CEO on a scale from -5 (*very unfavorable*) to 5 (*very favorable*) as in Study 4A. 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

Overall impressions. Overall impressions of the CEO differed by condition ($F(2, 297) = 9.31$, $p = .0001$). Specifically, relative to when he did not disclose weaknesses, overall impressions of the CEO were higher in both the *speaking weakness* condition ($M_{speaking\_weakness} = 2.88$, $SD = 1.71$; $M_{control} = 1.98$, $SD = 2.07$), $t(197) = 3.34$, $p = .001$, Cohen's $d = 0.47$, and the *skills weakness* condition ($M_{skills\_weakness} = 3.04$, $SD = 1.82$; $M_{control} = 1.98$, $SD = 2.07$), $t(199) = 3.85$, $p < .001$, Cohen's $d = 0.54$. Overall impressions did not differ between the *speaking* and *skills weakness* conditions ($t(198) = .61$, $p = .54$, Cohen's $d = 0.09$).

Willingness to work. Willingness to work with the leader differed 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 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, Cohen's d = 0.41, and showed a directional but non-significant result in the *skills weakness* condition ($M_{skills\_weakness}$ = 5.76, SD = 1.28; $M_{control}$ = 5.43, SD = 1.40), t(199) = 1.85, p = .065, Cohen's d = 0.25. Willingness to work with the CEO did not differ between the *speaking* and *skills weakness* conditions (t(198) = 1.04, p = .30, Cohen's d = 0.16).

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: the index of mediation excluded zero (b = .782, SE = .221, 95% CI = [.381, 1.247]), suggesting a significant indirect effect (Hayes, 2017). We observed similar results when we used willingness to work with the leader as the dependent variable (b = .413, SE = .132, 95% CI = [.193, .712]). The direction of the mediations indicates that revealing a weakness led to greater perception of authenticity, which in turn contributed to better overall impression of the manager, and greater willingness to work with the manager. Controlling for competence, the mediation for both impression (b = .425, SE = .137, 95% CI = [.198, .743]) and wiliness to work (b = .194, SE = .071, 95% CI = [.082, .367]) held. Controlling for both competence and liking, the mediation for both impression (b = .238, SE = .096, 95% CI = [.081, .459]) and willingness to work (b = .109, SE = .054, 95% CI = [.029, .245]) also held. Perceived authenticity explained 58.2% of the variance of overall impressions and 38.1% of the variance of willingness to work.

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: the index of mediation

excluded zero ($b = .477$, $SE = .112$, 95% CI = [.273, .714]), suggesting a significant indirect

effect (Hayes, 2017). We observed similar results when we used willingness to work with the

leader as the dependent variable ($b = .249$, $SE = .071$, 95% CI = [.128, .405]). Controlling for

competence, the mediation for both impression ($b = .302$, $SE = .078$, 95% CI = [.173, .483]) and

wiliness to work ($b = .111$, $SE = .044$, 95% CI = [.046, .221]) held. Controlling for both

competence and liking, the mediation for both impression ($b = .210$, $SE = .055$, 95% CI = [.121,

.347]) and willingness to work ($b = .061$, $SE = .034$, 95% CI = [.010, .146]) also held. Perceived

authenticity explained 58.4% of the variance of overall impressions and 32.8% of the variance of

willingness to work.

Supplemental Study 1D: Female Disclosers

Results

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

($F(1, 395) = 7.00$, $p = .008$), such that the manager left better overall impressions when s/he

disclosed a weakness relative to when s/he did not do so ($M_{experimental} = 3.15$, $SD = 1.67$; $M_{control} =$

2.68, $SD = 1.82$), $t(397) = 2.70$, $p = .007$, Cohen's $d = 0.27$. 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 an 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

($M_{experimental} = 5.77$, $SD = 1.24$; $M_{control} = 5.53$, $SD = 1.25$), $t(397) = 1.98$, $p = .048$, Cohen's $d =$

0.20. 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. A 10,000 sample bootstrap analysis (Model 4) showed that the index of mediation excluded zero (b = .603, SE = .135, 95% CI = [.341, .871]), suggesting a significant indirect effect. A mediation analysis with willingness to work as the dependent variable to test mediation by authenticity (Model 4) showed that the index of mediation excluded zero (b = .333, SE = .077, 95% CI = [.189, .492]), suggesting a significant indirect effect (Hayes, 2017). The direction of the mediations indicates that revealing a weakness led to greater perception of authenticity, which in turn contributed to better overall impression of the manager, and greater willingness to work with the manager. Controlling for competence, the mediation effect still held for impression (b = .376, SE = .089, 95% CI = [.232, .592]) and willingness to work (b = .170, SE = .056, 95% CI = [.084, .731]). Perceived authenticity explained 58.8% of the variance of overall impressions, and 58.3% of the variance of willingness to work.

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.

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

Rudman, L. A. (1998). Self-promotion as a risk factor for women: the costs and benefits of counterstereotypical impression management. *Journal of Personality and Social Psychology*, *74*(3), 629.