

LBS Research Online

[D Efron](#) and M Raj

Disclosing interpersonal conflicts of interest: revealing whom we like, but not whom we dislike

Article

This version is available in the LBS Research Online repository: <https://lbsresearch.london.edu/id/eprint/1660/>

[Efron, D](#) and Raj, M

(2021)

Disclosing interpersonal conflicts of interest: revealing whom we like, but not whom we dislike.

Organizational Behavior and Human Decision Processes, 164. pp. 68-85. ISSN 0749-5978

DOI: <https://doi.org/10.1016/j.obhdp.2021.01.005>

Elsevier

<https://www.sciencedirect.com/science/article/pii/...>

Users may download and/or print one copy of any article(s) in LBS Research Online for purposes of research and/or private study. Further distribution of the material, or use for any commercial gain, is not permitted.

Running Head: INTERPERSONAL CONFLICTS OF INTEREST

Disclosing Interpersonal Conflicts of Interest:
Revealing Whom We Like, But Not Whom We Dislike

Daniel A. Effron¹
Organisational Behaviour Subject Area
London Business School

Medha Raj
Management and Organizations Department
Marshall School of Business
University of Southern California

Author Note

We have no conflicts of interest to disclose.

Address correspondence to Daniel A. Effron, London Business School, Regent's Park,
London, NW1 4SA, United Kingdom, deffron@london.edu

We thank Ena Inesi, Selin Kesebir, and Krishna Savani for helpful feedback, and Zareen Choudhury and Sharmay Mitchell for research assistance.

Abstract

Imagine your boss asks you to evaluate the work performance of a coworker whom you happen to like or dislike for reasons unrelated to the performance. This situation poses an *interpersonal conflict of interest* because the fact that you like or dislike the coworker could undermine your professional obligation to offer an objective evaluation. We hypothesize that people are less likely to disclose conflicts of interest that involve disliking as opposed to liking, because they worry that disclosing dislike will make them look unsociable. Nine studies (four pre-registered) – examining hypothetical, actual, and lab-simulated workplace conflicts of interest – provide supportive evidence, and cast doubt on alternative explanations based on the motivations to maintain objectivity or to get away with bias. The findings demonstrate the importance of considering interpersonal dynamics in theorizing about advice-giving and conflicts of interest. We discuss implications for detecting conflicts of interest in organizations.

Abstract word count: 146

Keywords: conflict of interest; disclosure; advice; impression management; bias

Disclosing Interpersonal Conflicts of Interest:

Revealing Whom We Like, But Not Whom We Dislike

Imagine that, to inform a promotion decision, your boss asks you to write a performance review of your coworker. Your boss is unaware that you happen to know your coworker outside of work and like him for reasons completely unrelated to his performance. Now suppose instead that you *dislike* your coworker for reasons unrelated to his performance. In both situations, your personal feelings about the coworker have the potential to bias your evaluation of his work. Although you have a professional obligation to render an objective judgment, the motivation to help someone you like or to hinder someone you dislike gives you a personal interest in the promotion decision's outcome. In this way, both situations present you with a *conflict of interest*, defined as a tension between a professional obligation and a personal incentive (Cain, Loewenstein, & Moore, 2005; Sah, Loewenstein, & Cain, 2013).¹ In both situations, your boss might want to know about the conflict of interest so that she could make an informed decision about whether to discount your evaluation. We propose, however, that you would be less likely to tell your boss about the conflict of interest if it involved disliking your colleague.

The present research examines how people decide whether to disclose interpersonal feelings that pose conflicts of interest. Conflicts of interest are common in professional settings in which *advisees* request objective advice or judgments from *advisors* (e.g., Church & Kuang, 2009; Hwong, Sah, & Lehmann, 2017; Moore, Tetlock, Tanlu, & Bazerman, 2006). A conflict arises when advisors have an incentive not to be objective. The incentive is often financial, such as when a banker can earn commissions by selling clients products they do not need (G. Smith, 2012), but can also be non-financial (see Maj, 2008; Saver, 2012), such as when a peer-reviewer

¹ Tensions between competing professional obligations also pose conflicts of interest.

could help a friend by recommending her manuscript for publication, or harm a rival by recommending rejection (PLoS Medicine Editors, 2008). It is the *potential* for the incentive to interfere with an advisor's professional obligations that defines a conflict of interest, regardless of whether the incentive *actually* influences the advisor (R. Smith, 2006).

We introduce the term *interpersonal conflicts of interest* to describe situations in which one's professional obligation to objectively evaluate someone's performance could be undermined by the fact that one likes or dislikes that person. Our research examines interpersonal conflicts of interest involving three parties: an *advisor* whom an *advisee* asks for an objective judgment that is related to a *target person* whom the advisor likes or dislikes. For example, an employee (the advisor) could be asked by a boss, peer, subordinate, committee, or client (the advisee) to evaluate a report, idea, or other output by a coworker (the target person) whom the employee likes or dislikes.

An advisor could have many reasons for liking or disliking the target person. Perhaps the target has been a friend or an enemy for years, perhaps the target once helped or harmed the advisor, or perhaps the advisor has observed the target treating others kindly or rudely. Crucially, however, to constitute a conflict of interest, *the reason the advisor likes or dislikes the target person must be normatively irrelevant to the performance that the advisor must judge*. For example, imagine a software developer who dislikes her coworker because he was rude to her partner at the office party. The developer would face an interpersonal conflict of interest if she were asked to evaluate the quality of her coworker's computer code, because her reasons for disliking him have nothing to do with his coding performance. By contrast, she would *not* face an interpersonal conflict of interest if she were asked to evaluate the ease of collaborating with him, because her dislike for him would make such a collaboration objectively difficult. Some

situations will present more ambiguity about whether the advisors' liking or disliking are relevant to the advisors' performance. For example, organization members may disagree about whether promotion decisions should consider how well-liked a candidate is. The present research, however, avoids such ambiguity by focusing on situations in which the advisor likes or dislikes a target person for reasons completely unrelated to the target's performance.

Interpersonal feelings of all intensities can create conflicts of interest – for example, even mild liking or disliking for a person can bias judgments of their performance (e.g., Landy & Sigall, 1974). Analogously, financial incentives of all sizes can create conflicts of interest – for example, there is concern that even small gifts from drug companies have the potential to bias doctors' judgments (Dana, 2009; Dana & Loewenstein, 2003). Our research thus examines situations in which advisors may have strong or mild feelings of liking or disliking for target people. Our theorizing does not suggest any hypotheses about whether the strength of feelings moderates our results, but our studies still explore this issue.

We propose that advisors are less likely to disclose an interpersonal conflict of interest to advisees if the conflict involves disliking, as opposed to liking, the target person. The following sections develops this prediction by reviewing work on conflicts of interest and their disclosure, outlining two potential mechanisms that informed our prediction, and presenting five main studies testing this prediction and these mechanisms while addressing alternative explanations.

Significance and Literature Review

For several reasons, it is important to understand how advisors decide whether to disclose interpersonal conflicts of interest. First, a conflict of interest can bias advisors' judgment, even without their awareness, leading them to offer suboptimal advice (Dana & Loewenstein, 2003; Moore & Loewenstein, 2004). Such bias can arise not only from financial incentives (Moore,

Tanlu, & Bazerman, 2010; see also Babcock & Loewenstein, 1997; Moore et al., 2006; Messick & Sentis, 1979), but also from feelings of liking or disliking (Cardy & Dobbins, 1986; Dobbins & Russell, 1986; Feldman, 1981; Johnson, Erez, Kiker, & Motowidlo, 2002; Regan, Straus, & Fazio, 1974; Turban, Jones, & Rozelle, 1990). For example, students rated a professor's classroom performance less accurately when they were given an irrelevant reason to like or dislike the professor (Landy & Sigall, 1974). Understanding when advisors will conceal versus disclose their conflicts of interest can help predict the situations in which advisors' bias is likely to go undetected.

A second reason to study advisors' disclosure of interpersonal conflicts of interest is that such disclosures are likely to alter advisees' decisions. Research on financial conflicts of interest demonstrates both positive and negative consequences of disclosure (see Sah, 2019). On one hand, disclosures could improve advisees' decisions by closing the information gap between themselves and their advisors (Crawford & Sobel, 1982). In response to the disclosure, advisees could give less weight to the advisor's input (Church & Kuang, 2009), seek a second opinion (Schwartz, Luce, & Ariely, 2011), or ask the advisor to recuse him or herself from the decision at hand (Thompson, 1993). On the other hand, disclosure sometimes worsens advisees' decisions by making advisors feel licensed to offer more-biased advice (Cain et al., 2005; Cain, Loewenstein, & Moore, 2011; Koch & Schmidt, 2010), and by making advisees feel pressured to follow this advice without adequately discounting it or seeking a second opinion (Cain et al., 2005; Rose et al., 2019; Sah, Fagerlin, & Ubel, 2016; Sah et al., 2013; Schwartz et al., 2011). Because disclosure can have these consequences on decision-making, it is important to understand when advisors will choose to disclose.

A third reason to investigate this topic is that advisors may often have discretion about whether to disclose interpersonal conflicts of interest. Organizations' disclosure policies can become vague and difficult to enforce when they touch on interpersonal conflicts of interest. For example, *Organizational Behavior and Human Decision Processes* requires a peer reviewer to disclose any "personal relationships" that "could affect his/her objectivity, or inappropriately influence his her actions" (Elsevier, 2019, March). However, reviewers may not know what counts as a "personal relationship" (the line between friend and acquaintance is blurry); they can experience an interpersonal conflict of interest even in the absence of a formal relationship (e.g., they could dislike an author for making nasty comments at their job talk); and they will likely be overconfident in their ability to remain objective (Chugh, Bazerman, & Banaji, 2005). In sum, even in organizations with policies related to interpersonal conflicts of interest, members may still be able to choose whether to disclose such conflicts, these conflicts can compromise advisors' judgments, and disclosing conflicts of interest can affect advisees' decisions – for better or worse.

For these reasons, it is important to understand how advisors decide whether to disclose interpersonal conflicts of interest. Yet current theories about conflicts of interest are too narrow to shed much light on this question. Theories in accounting treat managers as rational agents who decide whether to disclose financial information to shareholders based on economic considerations, such as how disclosure will affect stock price (see Healy & Palepu, 2001). These theories do not intend to predict behavior in advisor-advisee relationships, nor do they consider the psychological considerations inherent in conflicts of interest.

Conflict-of-interest theories in organizational behavior do consider these dynamics, but do not consider why advisors would choose to disclose or conceal such conflicts. Perhaps

motivated by an interest in mandatory disclosure laws, the conflict-of-interest literature has focused on disclosure's consequences rather than its causes (e.g., Guo, Sriram, & Manchanda, 2019; Pham-Kanter, Alexander, & Nair, 2012). For example, one set of theories explains how disclosure affects the way advisors choose to balance financial self-interest with moral considerations (Cain et al., 2011), and how disclosure affects the social pressure advisees feel to comply with advice (Sah et al., 2013).

Another influential theoretical perspective in organizational behavior focuses on how conflicts of interest affect advisors' judgments. For example, the theories of *moral seduction* (Moore et al., 2006) and *bounded ethicality* (Chugh et al., 2005) both posit that advisors often succumb to conflicts of interest because they fail to appreciate their own susceptibility to bias. However, neither theory was formulated to predict when and why advisors will voluntarily disclose conflicts of interest.

None of these theories offers much insight about *interpersonal* conflicts of interest. Behavioral studies have manipulated conflicts of interest using financial incentives, and advisors in these studies typically offer advice about interpersonally inert stimuli, such as the value of coins in a jar (Cain et al., 2005) or the number of filled dots on a display (Sah & Loewenstein, 2014). As a result, these studies cannot speak to conflicts of interest that arise because advisors like or dislike the target person.

Like theories about conflicts of interest, theories about advice neglect the interpersonal dynamics that arise between advisors and advisees (Blunden & Gino, 2018), focusing instead on how advice affects accuracy (see Bonaccio & Dalal, 2006; Van Swol, Paik, & Prah, 2018). One exception is a theory positing that advisees accept advice not only to improve their accuracy, but also to reduce how responsible they feel for their decision, and to comply with social pressure

(Harvey & Fischer, 1997; see also Blunden, Logg, Brooks, John, & Gino, 2019). Although this theory considers interpersonal dynamics in advice-giving, it addresses neither interpersonal conflicts of interest nor why advisors would conceal or reveal them.

The present research addresses these limitations of previous theorizing about conflicts of interest and advice. First, by examining how people decide whether to reveal their conflicts of interest, we shed light on disclosure's causes, in contrast to prior theories' focus on its consequences. Second, by examining interpersonal conflicts of interest, we broaden the scope of prior theorizing beyond conflicts of interest that arise from financial incentives. Third, we posit that interpersonal concerns not only create conflicts of interest, but also influence advisors' behavior in such situations. As explained in the next section, we theorize that advisors' decisions to disclose whom they like versus dislike are driven, in large part, by a desire to be liked themselves.

By broadening the field's focus so that it encompasses interpersonal conflicts of interest, we highlight a novel research question. Are advisors less likely to disclose a conflict of interest to advisees if it stems from disliking (vs. liking) the target person? The next section develops our hypothesis that the answer is yes.

Willingness to Disclose Interpersonal Conflicts of Interest

Multiple motives could influence advisors' decisions about disclosing interpersonal conflicts of interest. We suggest that two motives in particular could lead advisors to be more willing to disclose whom they like than whom they dislike in the context of such conflicts: the desire to be liked, and the ethical obligation to ensure objectivity. We discuss each in turn.

The Desire to be Liked: Impression-Management Concerns About Sociability

We propose that impression-management concerns influence advisors' inclinations to disclose interpersonal conflicts of interest. According to impression-management theory, the desire to shape how people perceive us drives much of our public behavior (Leary, 1995), particularly in the workplace (Bolino, Long, & Turnley, 2016; Dubrin, 2011; Gardner & Martinko, 1988). This desire not only determines the behaviors we perform, but also the behaviors we avoid (Arkin, 1981; Ashford & Northcraft, 1992; Brooks, Gino, & Schweitzer, 2015). For example, employees will conceal personal information about themselves that they think will make them look bad (John, Barasz, & Norton, 2016).

A key impression that most people wish to make is that they are *sociable* (Jones & Pittman, 1982; Leary, 1995; Wayne & Ferris, 1990). One of the three fundamental dimensions on which people form impressions of others (along with competence and morality), sociability encompasses traits such as likeable, agreeable, warm, and friendly (Goodwin, 2015; Goodwin, Piazza, & Rozin, 2014; Leach, Ellemers, & Barreto, 2007).² The desire to manage an impression of sociability may influence advisors' disclosure decisions. Specifically, advisors may hold a lay theory that disclosing whom they dislike would make them appear dislikeable themselves, whereas disclosing whom they like would make them appear likeable.

Such a lay theory would be accurate. Audiences tend to dislike a person more when he or she evaluates others unfavorably versus favorably (Amabile, 1983; Ames, Bianchi, & Magee, 2010; Folkes & Sears, 1977; Gawronski & Walther, 2008; Wyer, Budesheim, & Lambert, 1990). One reason is that they infer that a person has the very traits he or she ascribes to others – a phenomenon called *spontaneous trait transference* (Carlston & Skowronski, 2005; Mae, Carlston, & Skowronski, 1999; Skowronski, Carlston, Mae, & Crawford, 1998). Thus,

² Morality and sociability are empirically distinct dimensions (Brambilla & Leach, 2014), even though earlier theorizing combined them into a single “warmth” dimension (Fiske, Cuddy, Glick, & Xu, 2002).

expressing dislike for someone can make you seem dislikeable (unsociable) yourself. Another reason is that social norms proscribe speaking poorly of others (Bergsieker, Leslie, Constantine, & Fiske, 2012), and violating social norms is usually considered unsociable. Thus, if advisors are concerned with how sociable they appear, they should be less likely to disclose whom they dislike than whom they like.

In theory, people could be reluctant to disclose whom they dislike to avoid appearing unsociable, eager to disclose whom they like in order to appear sociable, or both. However, the motivation to avoid making negative impressions is likely stronger than the motivation to make positive impressions (see Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Rozin & Royzman, 2001). Moreover, people may anticipate that violating a norm by speaking poorly of someone would create more of a negative impression than upholding a norm by speaking well of someone would create a positive impression (Reeder & Spores, 1983; Risky & Birnbaum, 1974). For these reasons, we emphasize a reluctance to disclose dislike instead of an eagerness to disclose liking. (Study 5 tests whether this emphasis is warranted).

Our theorizing focuses on impression-management concerns about sociability in particular because, according to the principle of spontaneous trait transference, disclosing that you like or dislike someone will affect impressions of your sociability (i.e., likeability) without necessarily affecting impressions of your other traits like morality or competence (Skowronski et al., 1998). Of course, advisors may also desire to manage impressions of other traits (see Jones & Pittman, 1982), and such desires may also influence disclosure decisions. For example, advisors may expect that disclosing any conflict of interest – whether financial or interpersonal – would make them look moral. However, it is less clear how the desire to appear moral or competent would make people more reluctant to disclose whom they dislike than whom they like. Our

research examines whether impression-management concerns about sociability, above and beyond concerns about other impressions, explain advisors' disclosure preferences.

In sum, the desire to be liked could make advisors prefer to disclose whom they like than whom they dislike. But there is also another motive that could create the same preference: an ethical obligation to ensure objectivity.

The Obligation to Ensure Objectivity: Bounded Ethicality and Concerns About Others' Bias

Conflicts of interest are ethically fraught situations because, by definition, they involve an incentive to violate a professional obligation to be objective (Sah et al., 2013). According to the theory of bounded ethicality, people care about being ethical, but cognitive biases can prevent them from recognizing when they are violating ethical principles (Bazerman & Tenbrunsel, 2012; Chugh et al., 2005; Chugh & Kern, 2016; see also Tenbrunsel & Messick, 2004). One such bias is overconfidence in one's own objectivity, known as the *bias blind spot* (Pronin, 2007; Pronin, Gilovich, & Ross, 2004; Pronin, Lin, & Ross, 2002). A bounded-ethicality analysis of conflicts of interest suggests that, because most advisors want to be ethical, they do not intend to allow such conflicts to compromise their judgments; rather, they overestimate their ability to remain objective in the face of such conflicts, and therefore fail to take steps to mitigate them (Chugh et al., 2005; see also Moore et al., 2006).

Extending bounded ethicality theory, we suggest that advisors will be more inclined to disclose an interpersonal conflict of interest if it meets their ethical obligation to uphold the principle of objectivity. However, it is not their own objectivity that may concern advisors, but rather the objectivity of their advisees. According to the bias blind spot, people think others are more susceptible to bias than themselves (Armor, 1999; Pronin, 2007). For example, people

readily acknowledge that interpersonal liking biases others' judgments, but fail to realize when their own judgments exhibit the same bias (McPherson Frantz, 2006). Thus, advisors may think that they can objectively evaluate a target person's performance despite their feelings about her, but worry that knowledge of those feelings would bias advisees' evaluations. That is, knowing that the target is liked could bias advisees in favor of the target, whereas knowing that the target is disliked could bias advisees against the target – much like inadmissible character evidence about a defendant biases juries' evaluations of a defendant's guilt (Hunt & Budesheim, 2004). Imagine being on a committee charged with selecting the best conference submission. Telling the committee that you hate one of the authors under consideration, for reasons completely unrelated to the quality of the submission, might bias the committee against the author. The committee could start disliking the author themselves or feel pressure not to give the person the award. In short, advisors could be reluctant to disclose interpersonal conflicts of interest that they think could bias the advisees' judgment.

This reluctance could make advisors more likely to conceal whom they dislike than whom they like. Advisors could anticipate that negative information about a target person influences others' judgment more than positive information – a phenomenon known as the *negativity bias* in person perception (see Skowronski & Carlston, 1989). Thus, they could anticipate that saying a target person is dislikeable would bias advisees' judgments more than saying the target person is likeable. Telling the best-paper committee that you love one of the authors might unfairly advantage her, but telling that committee that you hate one of the authors is more likely to unfairly disadvantage her.

To summarize, advisors may think that revealing whom they dislike would bias others' judgments more than revealing whom they like. Given that advisors feel ethically obligated to

uphold the principle of objectivity, they should thus be more reluctant to disclose whom they dislike than whom they like.

Hypotheses

Based on the above reasoning, we hypothesized that:

Hypothesis 1 (H1): Advisors will be less inclined to disclose an interpersonal conflict of interest to an advisee when the conflict involves disliking (vs. liking) a target person.

We formulated two hypotheses about mechanism, and were agnostic about whether one or both would explain this effect:

Hypothesis 2a (H2a): Impression-management concerns about sociability explain why advisors are less likely to disclose whom they dislike than whom they like.

Hypothesis 2b (H2b): Concerns about advisees' objectivity explain why advisors are less likely to disclose whom they dislike than whom they like.

Alternative Mechanism

A potential alternative mechanism is that advisors conceal interpersonal conflicts of interest to get away with offering intentionally biased advice – and that their advice is more likely to be biased against a target person they dislike than it is to be biased in favor of a target they like. Such a mechanism would be inconsistent with specific findings suggesting that people are less likely to harm those they have a reason to dislike than to help those they have a reason to like (Brewer, 1999; Buttelmann & Böhm, 2014; De Dreu et al., 2010; Halevy, Bornstein, & Sagiv, 2008; Halevy, Weisel, & Bornstein, 2012; Weisel & Böhm, 2015; Wang, Galinsky, & Murningham, 2009; Whitson, Wang, See, Baker, & Murnighan, 2015). Nonetheless, we address this potential mechanism in Study 2.

Overview of Studies

We present nine studies testing our hypotheses. Studies 1 and 2 tested whether advisors would be more inclined to disclose positive than negative interpersonal conflicts of interest (H1). Study 1 examined retrospective reports of disclosure decisions at work, whereas Study 2 observed real disclosure behavior in the lab. Studies 3–5 tested our hypotheses about mechanism (H2a and H2b) in vignette experiments using both measurement-of-mediation and moderation-of-process designs (Spencer, Zanna, & Fong, 2005). Study 5 added a control condition to explore whether people are reluctant to disclose whom they dislike, eager to disclose whom they like, or both. It also explored whether advisors' preferences for disclosing their feelings about a target person depended on how positively or negatively they evaluated the target's performance. Four additional studies in the Online Supplement (S1a–S1c and S2) provide additional evidence of robustness, generalizability, and mechanism.

To address alternative explanations, we also tested whether advisors are more likely to conceal whom they dislike (vs. like) because they think their advisee already knows whom they dislike (Studies 1 and 2), because they want get away with offering more-biased advice about people they dislike (Study 2), or because they want to manage impressions of their competence, credibility, or morality (Study 5). Across studies, we also assess whether the results generalize across different roles the advisee could occupy relative to the advisor: boss, peer, and subordinate.

Open Practices

We report all measures and conditions. In each study, we determined our stopping rules for data collection before running any participants. Verbatim materials for all studies and pre-registration documents for Studies 2, 4, 5, and S2 are available at

https://osf.io/npzw4/?view_only=e1fe082d5f54479fb3074492daf5da82, and data are available upon request from the first author.³

Study 1

Method

Participants. All 485 students in a Masters of Business Administration (MBA) course in the UK were invited to complete Study 1 as part of a longer questionnaire for their coursework. The class was 40% female, represented 64 nationalities, and had an average of 5.5 years of work experience (range = 2 to 15). A plurality of students had worked in consulting (28%) and finance (25%), with the remainder spread across a wide variety of industries. After dropping data from students who had previously submitted responses to the study or did not consent to participate in research, 431 people remained. As detailed in the Methods and Results sections, participants could only respond to the dependent measure if they could recall a relevant experience. We received 286 responses to the dependent measure from 225 participants in this repeated-measures design.

Statistical power. Calculations described in the Online Supplement show that Study 1 had enough statistical power to detect small-to-moderate condition differences. Specifically, if 25%, 50%, or 75% of people with a negative relationship said they had disclosed it, a between-participants design would have 80% power to detect at $\alpha = .05$ (two-tailed) that, respectively, 41%, 67%, or 88% of people with a positive relationship said they had disclosed it, given the cell sizes of our final sample.

Procedure. Study 1 was embedded in a longer questionnaire of unrelated measures for a class demonstration. The measures were presented in random orders. We randomly assigned

³ The Research Ethics Committee that approved this research withheld permission to post our data in a public repository because the consent forms did not inform participants that their responses might be posted.

participants to recall a coworker from a previous job with whom they had either had a strongly positive relationship (*liking condition*) or a strongly negative relationship (*disliking condition*) “for reasons unrelated to [the coworker’s] work performance.” In the liking condition, we explained the coworker could be someone “whom you’d go out of your way to talk with at work, or whom you enjoyed socializing with outside of work,” and in the disliking condition, we explained it could be someone “whom you’d go out of your way to avoid at work, or whom you would never want to socialize with outside of work.”

If participants could recall such a person, they recorded the coworker’s initials, reported how much they liked this person as a manipulation check ($-3 = \textit{strongly dislike}$, $3 = \textit{strongly like}$), and jogged their memory by writing 1–2 sentences explaining why their relationship with the coworker was positive or negative. Then they indicated whether they had ever been asked to evaluate coworker’s work performance (e.g., to fill out a performance review for HR, or to provide an informal opinion to a boss), and, if so, to describe the situation. The dependent measure then assessed disclosure: “Did you tell the person who asked you to evaluate [coworker’s initials] about your [positive/negative] relationship with [coworker’s initials]?” (yes/no).

As ancillary measures, we next asked “Did the person who asked you to evaluate [coworker’s initials] already know about your [positive/negative] relationship with [coworker’s initials]?”, and we requested the coworker’s gender. Finally, participants imagined that their boss at their next job is considering the same coworker for a position at the company, and asks participants to objectively evaluate the coworker’s prior work performance. As a second measure of the dependent variable, participants indicated how likely they would be to tell their boss about their relationship with the coworker ($-3 = \textit{extremely unlikely}$, $3 = \textit{extremely likely}$).

Finally, participants who had been randomly assigned to recall a positive relationship repeated the procedure while recalling a negative relationship, and vice versa. Thus, the study had a two-cell, within-subjects design (feelings about coworker: liking vs. disliking).

Results

Analytic strategy. To account for the fact that each participant could provide responses about up to two coworkers (one liked and one disliked), we analyzed the data using multi-level regressions with random intercepts for participants, and a dummy code for relationship type (1 = positive, 0 = negative).

Manipulation check. The majority of people (280, or 65% of $N = 431$) could recall a negative work relationship, and almost everyone (414, or 97% of $N = 431$) could recall a positive work relationship. Indicating that participants followed instructions, the manipulation check showed that people liked the coworker with whom they had a positive relationship ($M = 2.61$, $SD = 0.65$) more than the coworker with whom they had a negative relationship ($M = -1.68$, $SD = 1.06$), $d = 4.87$, $b = 4.28$, $SE(b) = 0.06$, $z = 66.25$, $p < .001$. The range of scores on this measure indicates that participants considered target people whom they strongly disliked (-3), strongly liked (3), and all scores in between.

Past disclosure behavior. Participants were only shown the primary dependent measure if they said they had been asked to evaluate their coworker's work performance. The proportion of participants who were shown this measure was similar among those who could recall a coworker they liked (42%, or 178 out of 414) as among those who could recall a coworker they disliked (39%, or 108 out of 280), $OR = 1.28$, $SE = .23$, $z = 1.38$, $p = .168$.

Consistent with H1, people were less likely to say they had disclosed their relationship with a coworker whom they disliked (63%, or 68 out of 108 responses) versus liked (75%, or

134 out of 178 responses). This difference was significant in a mixed logistic regression model (dependent measure coded 1 = disclosed, 0 = did not disclose; condition coded 1 = liking, 0 = disliking), $OR = 2.25$, $SE = 0.82$, $z = 2.22$, $p = .026$.

A possible explanation for this effect is that participants felt less need to disclose a negative relationship because they assumed that the advisee would already know whom they disliked. However, the data provided no support for this explanation. In fact, the reverse was the case: A greater proportion of participants said that the person who asked them to evaluate the coworker already knew about the relationship when it involved liking (76%, or 135 out of 178 responses) versus disliking (43%, or 46 out of 108 responses), $OR = 10.01$, $SE = 6.27$, $d = 0.71$, $z = 3.68$, $p < .001$. These findings therefore suggest that participants were more likely to disclose a relationship when disclosure was least informative: when advisees already knew about it.

Likelihood of future disclosure. Finally, we examined how likely participants thought they would be to disclose the relationship to a future boss who was thinking of hiring the coworker. We received responses to this measure from 99% of participants who could think of a positive relationship and 98% of participants who could think of a negative relationship. As predicted, participants said they would be less likely to disclose the negative relationship ($M = 0.89$, $SD = 1.84$) than the positive relationship ($M = 2.30$, $SD = 1.05$), $d = 0.94$, $b = 1.40$, $SE(b) = 0.10$, $z = 14.03$, $p < .001$.

Strength of liking and disliking. An exploratory analysis showed that our results did not significantly depend on how strongly participants liked or disliked their coworker. We operationalized feeling strength as the absolute value of participants' ratings on the manipulation-check measure of liking or disliking. Responses fell along the whole range from 0 (neutral feelings) to 3 (very strong feelings). We entered feeling strength and its interaction with

relationship type (positive or negative) into the regression analyses described previously.

Participants were more likely to report that they had disclosed the relationship when it was positive than when it was negative, regardless of how strongly they felt about the coworker, $OR = 1.21, z = .39, p = .699$ for the interaction. They also said that they would be more likely to disclose the positive than the negative relationship in the future, regardless of the strength of these feelings, $b = .04, SE(b) = .17, z = .27, p = .784$ for the interaction.

Discussion

MBA students' reports of how they had behaved in real-world work situations supported H1. When recalling a time when someone had asked them to evaluate a coworker's performance, participants said they were less likely to have disclosed an interpersonal conflict of interest if it arose from disliking a colleague than if it arose from liking a colleague. We found no evidence that this effect depended on whether their feelings towards the coworker were strong or mild. Strikingly, the effect also emerged even though disclosures of liking were less informative to the person requesting the evaluation (i.e., the advisee) than disclosures of disliking. That is, the advisee was more likely to already know about participants' relationship with the liked (vs. disliked) coworker. We also found that when participants imagined facing an interpersonal conflict of interest in the future, they thought they would be less likely to disclose that they disliked (vs. liked) the coworker.

Participants' retrospective reports offer evidence of external validity. However, we do not know whether these reports were accurate. Additionally, the fact that more people were able to recall negative versus positive work relationships could raise questions about selection effects. It is also possible that participants disclosed whom they liked precisely because the advisee already knew about these positively relationships. That is, participants may have wanted to strategically

signal their truthfulness by disclosing information that the advisee already knew to be true. Finally, because Study 1 was correlational, we cannot rule out the possibility that unobserved variables explain the results. To address these limitations, Study 2 brought participants into the lab where we could manipulate their feelings towards a target person, measure their actual disclosure behavior, and ensure that the advisee would have no way of knowing about participants' feelings towards the target person unless participants disclosed these feelings.

Study 2

Study 2 was a laboratory analogue to the real-world situations that Study 1 examined. Participants (advisors) were tasked with offering a “boss” (the advisee) an impartial evaluation of how well someone (the target person) had performed on a task, knowing that their evaluation could affect this target person's financial outcomes. We created an interpersonal conflict of interest by giving participants a reason, unrelated to the target person's performance, to like or dislike the target. Participants then chose whether to disclose their feelings about the target to the advisee. We predicted they would be less likely to disclose disliking than liking (H1).

We also examined whether this result could be explained by participants offering more-negative evaluations when they disliked the target. As noted, perhaps advisors conceal conflicts of interest to get away with offering intentionally biased advice about people they dislike. Conversely, our hypothesized mechanisms (H2a and H2b) predict that advisors would still be more reluctant to disclose negative than positive feelings even after accounting for the specific advice they offer.

Method

We pre-registered this study at <https://aspredicted.org/rd9cq.pdf>.

Participants. Residents of a major UK city who were enrolled in a behavioral lab's subject pool received £10 to spend about 1h completing Study 2 in the lab as part of a series of six unrelated studies, grouped together for efficiency and presented in random orders. We aimed to recruit 300 participants, based on the resources available and the sample sizes required for the unrelated studies being run simultaneously. Because we overbooked time slots to guard against no-shows, we received data from 309 participants. As detailed below, we included four questions to check comprehension of the instructions. If participants answered any incorrectly, they were shown the directions again and given another chance. Following our pre-registered exclusion criteria, we discarded data from the 22 people who answered incorrectly on their second chance, leaving 287 for analysis (146 in the disliking condition and 141 in the liking condition). Of the people who provided demographics, there were 202 women, 81 men, and 1 non-binary person; 181 students and 102 non-students; 112 current employees and 171 non-employees ($M_{\text{age}} = 26$ years, $SD = 9$). Sixty-four different nationalities were represented.

Statistical power. A sensitivity analysis described in the Online Supplement suggests that this study had enough statistical power to detect small effects. Specifically, if 25%, 50%, or 75% of people chose to disclose negative feelings, then Study 2 would have 80% power to detect that, respectively 32%, 57%, and 81% of people disclosed positive feelings

Method

Overview. Participants completed this study on computers in private lab rooms. We first manipulated whether they liked or disliked a target person (referred to as “the manager”). Then we asked them to evaluate the target's performance on an unrelated task, and to send the evaluation, along with advice about how much pay the target's performance merited, to an advisee (referred to as “the manager's boss”) – knowing that the advisee would (ostensibly)

determine how much money the target person received. Finally, we measured whether participants disclosed their feelings about the target person to the advisee. The next sections provide more detail.

Cover story and manipulation. As a cover story, participants read about a previous study ostensibly run in our lab (actually fictitious) in which people randomly assigned to the role of “employees” wrote an essay about a meaningful event in their lives, and people randomly assigned to the role of “managers” gave the employees written feedback on the essay. Study 2’s participants learned that their job would be to rate the feedback from one of these managers.

Once participants had read this information, a female experimenter entered the room with a stack of packets, each of which ostensibly contained a different employee’s essay and the corresponding manager’s feedback. The experimenter fanned out the packets face-down and participants chose one. In reality, there were only two versions of the packet, and participants’ choice randomly assigned them to experimental condition while keeping the experimenter blind.

In both conditions, the employee’s personal essay described moving to participants’ city, struggling to make friends and “fit in,” worrying about being perceived as “strange” due to cultural differences, and feeling “so lonely” but also “proud of being here on my own and making it work.” After reading the employee’s essay, participants read the manager’s feedback; to increase realism, the essay and the feedback were written in different handwriting with different colored pens.

We varied the manager’s feedback in order to influence participants’ feelings about the manager. In the *liking* condition, the manager gave the employee’s essay a high score (10/10), described the essay as “thoughtful and interesting,” and communicated sympathy and kindness: “Loneliness is no fun at all, and it’s hard feeling different. If you’d like to meet some cool people

in [city name], I'd love to introduce you to my friends, we're going to the pub tonight. Come say hi after the study if you'd like." In the *disliking* condition, the manager instead gave a low score (2/10), described the essay as "boring" and "not very thoughtful," and made rude, dismissive comments: "It's like a bad movie script. Feeling lonely and different doesn't make you special. Get over yourself."

Participants completed the remainder of the study on a computer. First, they entered the employee and manager's ID numbers shown on their packet. Doing so lent credence to the cover story that there were many different employees and managers that participants could have rated. It also recorded participants' experimental condition, because the manager had a different ID number in each condition. To further support the cover story, we asked participants to answer three filler items about the manager's feedback (how constructive, helpful, and accurate the feedback had been).

Manipulation check. To test whether the manager's feedback affected how much participants liked the manager, we asked, "How *positive* or *negative* is your impression of the manager" and "How much do you *like* or *dislike* the manager? (-3 = *Strongly [negative/dislike]*, +3 = *Strongly [positive/like]*, with intermediate points appropriately labelled). Responses to the two questions were highly correlated, $r = .96$, so we averaged them together.

Conflict of interest. Participants next learned about a task in which their feelings towards the manager would create an interpersonal conflict of interest. They were told that at the end of the previous study, the same manager had been asked to draw a space alien for a chance to earn a bonus gift voucher worth up to £10. The more creative the alien, the larger the manager's bonus. Two people would examine the drawing: the "manager's boss" (i.e., the advisee) who would determine the size of the manager's bonus, and the "advisor," who would advise the

manager's boss what the bonus should be. Participants learned that they would be the advisor, and that their advice would be sent to a future participant who would be assigned to the role of the manager's boss. A diagram reiterated the different roles and summarized the instructions (see Figure 1). After answering comprehension-check questions (described below), participants read the following instructions:

Important: Your ratings of the manager's alien drawing should be based only on its creativity – NOT on the feedback the manager gave the employee earlier in the study.

You must objectively evaluate the manager's creativity, based only on the drawing of the alien. It is important that you do not let any personal feelings about the manager influence your evaluation.

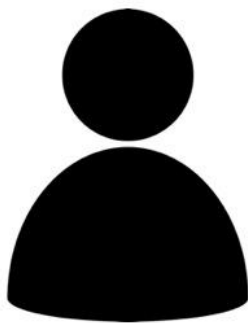
Thus, participants faced an interpersonal conflict of interest. On one hand, they were charged with sending an objective evaluation of a target person's work to an advisee. On the other hand, we expected the manipulation to give them positive or negative feelings about the target person that, though unrelated to the quality of the work, could influence how they evaluated it.

Dependent measure: disclosure behavior. We next showed participants their previous ratings of how much they liked the manager and how positive their impression was of the manager (i.e., their responses to the manipulation checks). Then we asked participants if they would like to share these ratings with the manager's boss (i.e., the advisee, who would be making the final decision about manager's bonus payment). The manager's boss would ostensibly receive the ratings at the same time as he or she received participants' evaluations of the alien drawing. Our dependent measure was whether participants disclosed their feelings about the manager by clicking "yes."

Figure 1

Diagram Shown to Study 2's Participants

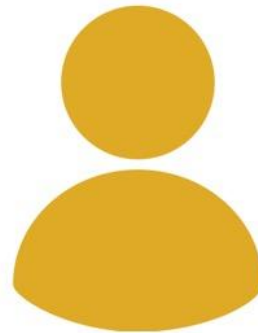
Please examine this diagram, which summarizes what you just read:

**Manager**

- Participant in previous study session
- Wrote the feedback you just read
- Drew an alien to receive a bonus payment

**Advisor**

- This is your role
- Rate how creative the manager's drawing is
- Rate how much of a bonus payment the manager should get

**Manager's Boss**

- Participant in future study session
- Will review the manager's alien drawing and your ratings
- Will make final decision about how much of a bonus the manager will receive

Exploratory measure: performance evaluation. We also included a behavioral measure of the advice participants actually sent to the manager's boss. Participants viewed a drawing of an alien, rated its creativity (7), and indicated how much of a bonus they thought the manager should receive "based only on the drawing" (£0 to £10 in £1 increments).

Comprehension checks. We asked three multiple-choice comprehension-check questions just before the dependent measure to ensure that participants understood the directions: What will you do in a moment? (*Answer: Rate a drawing*), Who will you give advice to? (*Answer: The manager's boss*), and What will be the consequence of your advice? (*Answer: It could affect the manager's bonus*). As noted, any participants who answered incorrectly were asked to review the directions, received a second chance to answer the questions, and were dropped from analysis if they still got any wrong.

Suspicion checks. A funneled debriefing procedure probed for suspicion. At the end of the study, participants were invited to type any comments they had about the study and what they thought the researchers were hoping to find. Then, after they had completed all six unrelated studies in the lab session, we gave them a brief description of each study and asked if they "were suspicious that anything might not be what it seemed." If they clicked "yes," we asked which study provoked suspicion, and what specifically they were suspicious about. Three people guessed that the research concerned how personal feelings affect disclosure decisions, and an additional 19 people suspected that the manager, employee, or survey materials were bogus. We retained these 22 participants for analysis, but excluding them did not change the results' direction or statistical significance.

All participants received a full debriefing at the study's conclusion.

Results

We pre-registered and report one-tailed tests for all confirmatory, directional predictions (Cho & Abe, 2013), but for simplicity we report two-tailed tests because they produced identical conclusions.

Manipulation check. Participants liked the target person (i.e., the “manager”) more in the liking condition ($M = 1.90$, $SD = 1.27$) than in the disliking condition ($M = -2.21$, $SD = .89$), $t(285) = 31.88$, $p < .001$, $d = 3.78$.

Disclosure behavior. As H1 predicted, fewer participants disclosed their feelings about the target person (“manager”) to the advisee (“manager’s boss”) when they disliked the target (54.11%) than when they liked the target (71.63%). This difference was statistically significant in a logistic regression model with a dummy code for condition (1 = liking, 0 = disliking), $OR = 2.14$, $SE = .54$, $z = 3.05$, $p = .002$.

Performance evaluation. Exploratory analyses suggested that participants’ feelings about the target person influenced the evaluations they sent to the advisee, despite their instructions to ignore these feelings. Specifically, participants rated the target person’s alien drawing as less creative in the disliking condition ($M = 5.86$, $SD = 1.79$) than in the liking condition ($M = 7.45$, $SD = 1.66$), $t(285) = 7.81$, $p < .001$, $d = .93$ – and, accordingly, they also advised a lower financial bonus for the target person in the disliking condition ($M = £5.37$, $SD = 2.11$) than in the liking condition ($M = £7.41$, $SD = 1.99$), $t(285) = 8.41$, $p < .001$, $d = 1.00$.

However, the fact that participants evaluated the target’s performance more negatively when they disliked the target cannot fully explain why participants were more reluctant to disclose disliking than liking. They were still less likely to disclose disliking than liking when we statistically controlled for their creativity ratings, $OR = 1.80$, $z = 2.15$, $p = .003$, and their advice about the bonus, $OR = 2.02$, $z = 2.52$, $p = .012$. (We controlled for creativity and advice in

separate analyses because these two measures were colinear, $r = .83, p < .001$). In fact, neither covariate significantly predicted disclosure in these analyses, respectively, $ORs = 1.12$ and 1.03 , $zs = 1.57$ and $.50$, $ps = .115$ and $.619$. In other words, even if participants in both conditions had sent equally favorable evaluations of the alien drawing to the advisee, they would still have been less likely to disclose that they disliked (vs. liked) the target person.

Additional exploration found no evidence that participants' relative reluctance to disclose dislike depended on how negatively they evaluated the target's performance. Specifically, neither the ratings of the alien drawing nor the advice about the bonus significantly interacted with the manipulation, $ORs = .83$ and 1.05 , $zs = 1.29$ and $.37$, $ps = .197$ and $.709$, respectively.

Strength of liking and disliking. An exploratory analysis found no evidence that our main results depended on the strength of participants' feelings about the target person. As in the previous study, we operationalized feeling strength as the absolute value of the manipulation-check measure of liking/positivity; responses fell along the entire possible range from 0 = neutral to 3 = very strong. When we entered this measure and its interaction with the manipulation into the logistic regression analysis, the interaction term was not significant, $OR = 1.10$, $z = .30$, $p = .762$.

Discussion

Study 2's participants faced a conflict of interest in the lab: On one hand, they were instructed to send an objective evaluation of a target person's work performance to an advisee, knowing that their evaluation could affect the target's financial compensation. On the other hand, they had a reason – unrelated to the target's performance – for either liking or disliking the target. The results provided behavioral evidence for H1: Participants were less likely to disclose their feelings towards the target to the advisee when we randomly assigned them to dislike (vs.

like) the target. Thus, the behavior we observed in Study 2 mirrored the behavior that participants recalled in Study 1. And as in Study 1, this effect emerged regardless of the strength of participants' liking or disliking for the target person. Addressing a limitation of Study 1, the effect also emerged under controlled conditions in which, absent participants' disclosure, the advisee had no way of knowing how participants felt about the target.

Why were participants more likely to disclose negative than positive feelings? The possibility that advisors conceal conflicts of interest to get away with offering biased advice about people they dislike was not a sufficient explanation. Participants were still less likely to disclose negative than positive feelings about the target person even after we statistically controlled for the size of the bonus they advised. Thus, we need to consider explanations that do not depend on the specific advice offered.

We have theorized that people think disclosing negative feelings will make them look unsociable (H2a) and/or bias others' judgments more than disclosing positive feelings (H2b). For example, Study 2's participants may have worried that revealing that they disliked the target person, more than revealing that they liked the target, would lead the advisee to infer that participants themselves were dislikable, or compromise the advisee's ability to objectively evaluate the target's work performance. Studies 3–5 test these mechanisms.

Study 3

Study 3 used a measurement-of-mediation design to test concerns about impression management and the advisee's bias as potential mechanisms. Participants indicated how likely they would be to disclose each of several interpersonal conflicts of interest in the workplace.

Method

Participants. We posted slots for 100 people on Prolific Academic, an online platform with participants who tend to be more naïve to experimental procedures and provide higher-quality data than MTurk participants (Peer, Brandimarte, Samat, & Acquisti, 2017). By *a priori* decision, we requested only participants from the UK, Ireland, and Canada, and we excluded any observations with IP addresses outside of the targeted countries, duplicate IP addresses, or duplicate participant IDs. The final sample contained 96 people (72 women, 24 men; $M_{\text{age}} = 35$ years, $SD = 12$; 49% employed full-time, 28% employed part-time).

Statistical power. Using the PANGEA web app (Westfall, 2016), we estimated that the smallest effect-size Study 3 could detect with 80% power and two-tailed $\alpha = .05$ was $d = .44$, which approximates the average published effect size in social psychology (Richard, Bond, & Stokes-Zoota, 2003).

Procedure. Participants read two vignettes, presented in randomized orders: one in which a boss asks them to write a performance evaluation of a coworker, and one in which participants are one of several judges deciding whether their coworker should win a product-naming contest (see Appendix). Depending on random assignment, participants had either a “strongly positive” or a “strongly negative” relationship with the coworker (i.e., the liking and disliking conditions, respectively). To emphasize that the vignettes involved interpersonal conflicts of interest, we told participants that their feelings about the coworker were irrelevant to the evaluation they had to make (e.g., “you have a strongly [positive/negative] relationship with this person, for reasons unrelated to their work performance”). We were interested in participants’ inclination to disclose this relationship to the advisees who were counting on participants’ objectivity (i.e., the boss or the other judges).

Measures. We randomized the order in which we measured the first two measures listed below.

Impression-management concerns about sociability. To test one potential mechanism, we asked participants to estimate how others would view them if they knew about participants' relationship with the coworker by selecting a response on each of four 7-point, bipolar scales, coded -3 to +3: cold/warm, unsociable/sociable, agreeable/disagreeable (reversed), rigid/easygoing (adapted from Goodwin, 2015). We averaged the items into a measure of sociability ($\alpha s > .91$ in each vignette).

Concerns about others' objectivity. To test another potential mechanism, we asked participants to predict how they would affect advisees' objectivity if they disclosed their relationship with the coworker. Specifically, we created this measure by averaging two highly correlated items: "If you told your boss about your relationship with this person ..." (1) "... how difficult would it be for the boss to remain objective when evaluating this person's work performance?" and (2) "... how much would this affect the boss's judgment about this person's work performance?" (*not at all, slightly, moderately, very much, completely*, coded 1–5), $r = .80, p < .001$. (In the vignette about the contest, "the boss" was replaced with "the other judges").

Main dependent measure: Likelihood of disclosure. We asked how likely people would be to disclose their relationship to the relevant advisees (i.e., the boss who had asked for the performance evaluation, or the other contest judges, depending on the vignette). Responses to this item ranged from 1 (*not at all likely*) to 5 (*definitely*).⁴

⁴ Study 3 included four additional exploratory measures: how concerned participants were about their *own* objectivity, how important participants thought disclosure was, how important they thought it was to recuse themselves from being an advisor, and how likely they would be to recuse themselves. See Online Supplement for details and results.

Results

Analytic approach. We submitted each measure to a mixed regression model with fixed effects for condition (1 = liking, 0 = disliking) and vignette, and random intercepts for participants.

Likelihood of disclosure (main dependent measure). Study 3 replicated our central finding. Supporting H1, people thought they would be less likely to disclose interpersonal conflicts of interest that involved disliking a colleague ($M = 2.23$, $SD = 1.20$) than those that involved liking a colleague ($M = 2.89$, $SD = 1.30$), $b = .65$, $SE(b) = .22$, $z = 3.00$, $p = .003$, $d = .53$.

Impression-management concerns. Consistent with our impression-management account of these findings (H2a), participants thought that their advisees would find them less sociable if they disclosed that they disliked a colleague ($M = -.59$, $SD = .98$) than if they disclosed that they liked a colleague ($M = .98$, $SD = .99$), $b = 1.57$, $SE(b) = .18$, $z = 8.52$, $p < .001$, $d = 1.59$.

Concerns about others' objectivity. Failing to support H2b, the results were not consistent with a mechanism based on concerns about others' objectivity. Participants did not think that they would bias their advisees' judgment more by disclosing that they disliked a colleague ($M = 2.03$, $SD = .81$) than by disclosing that they liked a colleague ($M = 2.18$, $SD = .99$). The means were in the opposite direction than predicted, though not significantly so, $b = .15$, $SE(b) = .16$, $z = .93$, $p = .350$, $d = .17$.

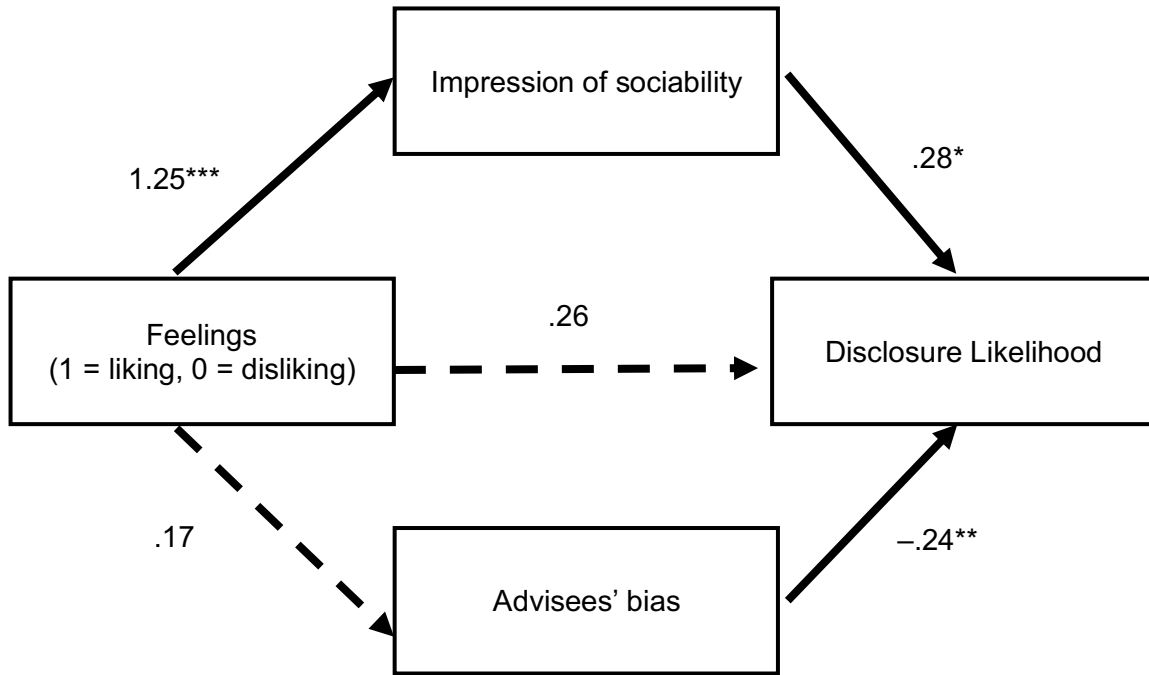
Mediation analysis. We next computed a parallel mediation model to test whether concerns about impression management and/or advisors' objectivity could explain participants' disclosure preferences. We used Stata's *gsem* command to estimate a series of mixed regression

models with dummy codes for condition (1 = positive relationship, 0 = negative relationship), z -scored mediators (estimates about appearing warm and beliefs about others' bias), and fixed effects for vignette and random intercepts for participants. Then we used the *nlcom* command to compute indirect effects by multiplying the relevant a and b paths together. This model tests each potential mediator while statistically controlling for the other. Figure 2 displays the results.

Consistent with the idea that impression-management concerns explain people's disclosure preferences (H2a), the results showed a significant indirect effect whereby people thought that disclosing liking would make them appear more sociable than disclosing disliking – and that the more sociable people thought they would appear, the more inclined they were to disclose, $b = .35$, $SE(b) = .16$, $z = 2.22$, $p = .026$. Inconsistent with the idea that concerns about others' objectivity explain people's disclosure preferences (H2b), beliefs about how much disclosure would bias advisees' judgments was not a significant mediator, $b = .04$, $SE(b) = .16$, $z = 1.04$, $p = .300$.

Figure 2

Impression-Management Concerns About Sociability Mediate the Reluctance to Disclose Dislike in Study 3



Note. Solid lines indicate statistically significant effects. Mediators were z-scored; disclosure likelihood was measured on a 5-point scale. * $p < .05$, ** $p < .01$, *** $p < .001$

Discussion

In a new vignette paradigm, Study 3 again found that advisors were less inclined to disclose whom they disliked than to disclose whom they liked (H1). Studies S1a–S1c in the Online Supplement provide evidence that this effect is robust and generalizable across different vignettes and populations (i.e., business students as well as online participants). The results of Study 3’s mediation analysis were consistent with the possibility that impression-management concerns about sociability could explain this effect (H2a). Advisors thought that disclosing dislike would make them appear less likeable to advisees than disclosing liking. Study S2 in the Online Supplement provides a pre-registered replication of this mediation effect using a different measure of impression-management concerns. We found no evidence for the possibility that concerns about others’ bias could explain the effect (H2b).

Study 4

Although Study 3’s data were consistent with an impression-management mechanism, no measurement-of-mediation analysis can demonstrate a causal mechanism (Bullock, Green, & Ha, 2010; Fiedler, Schott, & Meiser, 2011). To obtain stronger causal evidence for an impression-management mechanism, Study 4 experimentally manipulated these concerns in a moderation-of-process design (Spencer et al., 2005). Participants in this vignette study imagined that they either had strong or weak impression-management concerns. We expected people in the strong-concern condition to be less inclined to disclose an interpersonal conflict of interest if it arose from disliking as opposed to liking. However, we also expected this effect to attenuate in the weak-concern condition.

Method

We pre-registered Study 4 at <https://aspredicted.org/j5js7.pdf>.

Participants. All 497 MBA students taking a required introductory management class at an internationally diverse UK business school were invited to complete Study 4 as part of a longer class survey of unrelated measures (66 nationalities; 38% women and 62% men; 5 years of work experience on average). Of the 425 students who consented to participate, 421 gave a response to the dependent measure.

Statistical power. A sensitivity analysis conducted with G*Power showed that the smallest effect our final sample could detect at $\alpha = .05$ (one-tailed, as pre-registered) and 80% power was $f = .12$ for the predicted interaction term.

Procedure. The paradigm was adapted from Study 3. Participants read the *contest* scenario (see Appendix) under one of four randomly assigned conditions in a 2 (feelings towards target person: liking vs. disliking) vs. 2 (impression-management concerns: strong vs. weak) factorial design. Specifically, participants imagined being one of eight judges charged with picking the best idea to win a cash prize, and that the other judges are unaware that one of the finalists is someone with whom participants have a strongly positive (vs. negative) relationship for reasons unrelated to the quality of the idea. In the *strong impression-management concerns* condition, they further read (emphasis in original):

You are highly motivated to make a good impression on the other judges. You care a lot about what they think of you, and you want to present yourself to them as positively as possible.

In the *weak impression-management concerns* condition, they instead read:

You don't care at all about what the other judges think of you. You aren't really motivated to make a good impression on them, and it's not important to you to present yourself in any particular way to them.

Then participants indicated how likely they would be to tell the other judges about their relationship with the finalist (from 1 = *not at all likely* to 5 = *definitely*). Due to time limitations, we did not administer other vignettes or measures.

Results and Discussion

Because we had directional predictions, we pre-registered one-tailed significance tests, but we report two-tailed tests because they produced identical conclusions. We first tested main effects in a regression model with dummy codes for each independent variable, and then we added the interaction term in a second model (see Table 1).

A main effect of the feelings manipulation replicated our previous findings: Supporting H1, participants said they would be less likely to disclose the interpersonal conflict of interest in the disliking condition ($M = 2.00$, $SD = 1.14$) than in the liking condition ($M = 3.15$, $SD = 1.26$), $t(418) = 9.74$, $p < .001$, $d = .96$. This effect was qualified by a significant interaction with the impression-management manipulation, $b = .47$, $t(417) = 2.03$, $p = .043$ ($p = .022$ by the pre-registered one-tailed test), $f = .10$ (regression coding: liking condition = 1, disliking condition = 0; high impression-management = 1, low impression-management = 0). The positive interaction coefficient supports our pre-registered prediction, indicating that the effect of the feelings manipulation was stronger when impression-management concerns were high (vs. low; see Figure 3). Specifically, simple slopes analysis showed that when impression-management concerns were high, people were less inclined to disclose whom they disliked ($M = 1.93$, $SD = 1.14$) than whom they liked ($M = 3.31$, $SD = 1.21$), $b = 1.38$, $t(417) = 8.32$, $p < .001$, $d = 1.17$ – and when impression-management concerns were low, this effect was attenuated ($M_s = 2.08$ and 2.99 , $SD_s = 1.14$ and 1.30 , respectively) – though still significant, $b = .91$, $t(417) = 5.53$, $p < .001$, $d = .74$. That the effect remained significant despite its attenuation could suggest that

participants had other reasons in addition to impression-management for concealing whom they disliked. Another possibility is that the manipulation reduced but did not completely eliminate impression-management concerns.

These results provide experimental evidence that impression-management concerns help explain why people are less likely to disclose interpersonal conflicts of interest that involve disliking versus liking (H2a). Participants' greater reluctance to disclose whom they disliked than whom they liked was attenuated when participants imagined not caring what the advisees thought of them.

Table 1

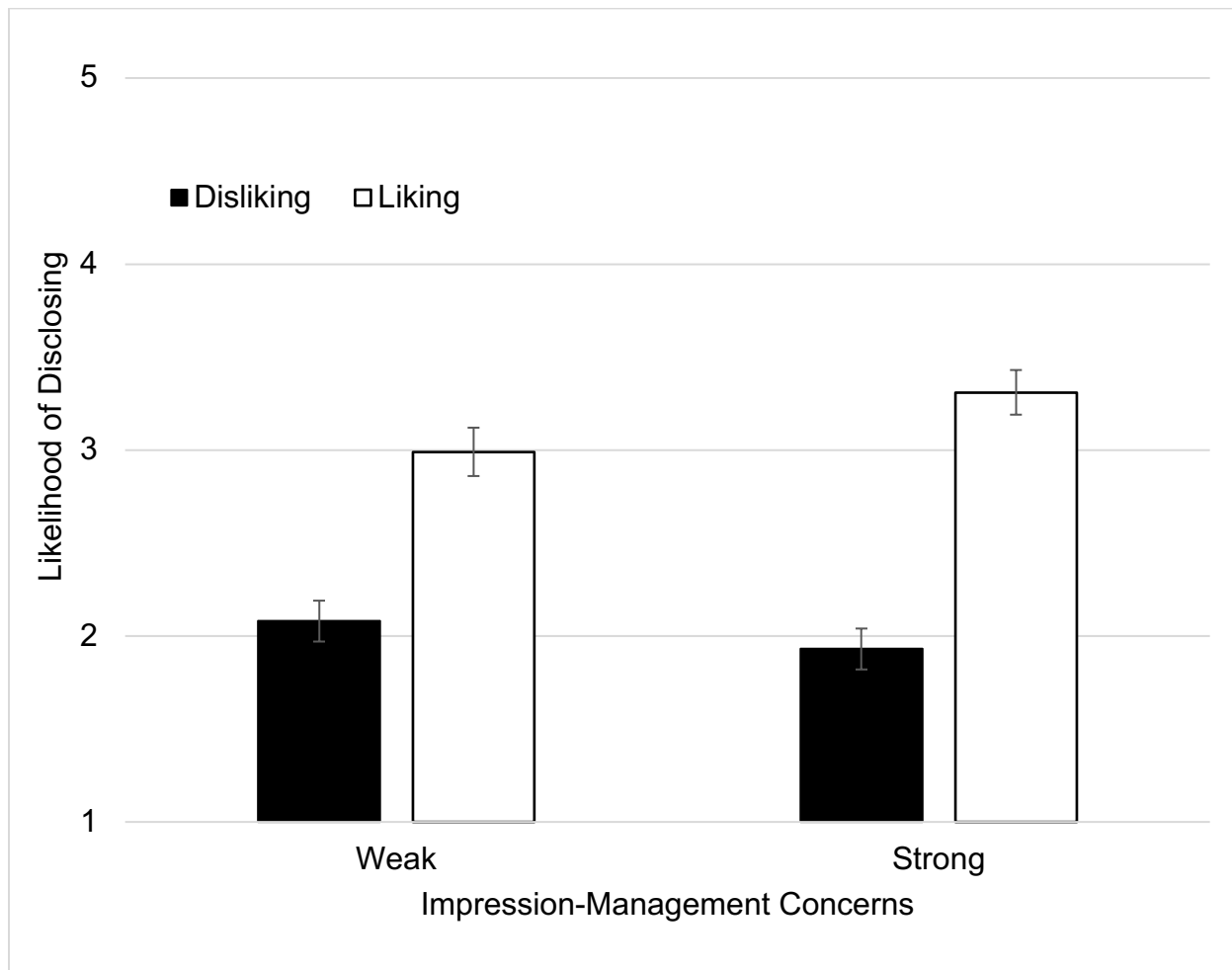
Study 4's Regression Results

	Main-Effect Model	Interaction Model
Liking condition	1.14*** (.12)	.91*** (.16)
Strong impression- management motives	.07 (.12)	-.16 (.16)
Interaction		.47* (.23)
Constant	1.97*** (.10)	2.08*** (.11)

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

Figure 3

Likelihood of disclosing liking vs. disliking for target person, as a function of experimentally manipulated impression-management concerns ($M \pm SE$).



Study 5

Having established clear evidence of the hypothesized phenomenon and provided evidence for an impression-management mechanism with both moderation and mediation, we conducted Study 5 to address a few remaining questions. First, by introducing a neutral-feelings control condition, Study 5 explored whether people are reluctant to disclose disliking, eager to disclose liking, or both.

Second, Study 5 tested whether people's reluctance to disclose disliking could be explained by impression-management concerns about sociability above and beyond impression-management concerns about competence and morality, the two other fundamental dimensions on which people form impressions (Brambilla & Leach, 2014; Goodwin, 2015). We expected to replicate Study 3's finding that concerns about sociability mediated the effect of disliking (vs. liking) on disclosure intentions. Going further, this finding should remain reliable even when accounting for alternative mediation pathways through concerns about competence and morality. We did not formulate predictions about whether these alternative pathways would be significant because, as noted, we had no theoretical basis for expecting people to think that disclosing dislike would make them appear less competence or moral than disclosing liking.

Third, we wanted to again test our hypothesis that concerns about advisees' objectivity help explain advisors' disclosure preferences (H2b). We found no support for this hypothesis in Study 3, but perhaps the specific measure of objectivity concerns was not sensitive enough to detect a real effect. Study 5 thus included a new measure of objectivity concerns.

Fourth, we explored whether a desire to appear credible could play a role in disclosure decisions. Advisors might expect to appear more credible if they disclosed interpersonal feelings that conflicted with their evaluations of the target person's performance. For example, describing

a coworker's work performance as excellent should seem more credible if you have disclosed that you dislike the coworker. Thus, if advisors want to appear credible, they should be most inclined to disclose interpersonal conflicts of interest when their evaluations and interpersonal feelings are misaligned. Study 2 found no evidence of this possibility: Lab participants were more likely to disclose that they disliked versus liked the target person, regardless of how they evaluated the target's work performance. Nonetheless, Study 5 provided an additional test by manipulating whether participants imagined offering a positive or negative evaluation.

Finally, Study 5 tested whether our findings would generalize across different types of relationships between the advisor and the advisee. The advisees in our previous studies were the advisor's boss, peers, or from their company's human resources division; Study 5 included a vignette in which the advisees were the advisor's subordinates.

Method

We pre-registered this study at <https://aspredicted.org/ze7nv.pdf>.

Participants. Study 5's participants were employees with managerial experience, recruited from Prolific Academic. We requested 600 complete responses from Prolific, because an *a priori* power analysis with the PANGEA web app suggested that a sample this size would provide 80.5% power to detect a small interaction effect, $d = .20$, in Study 5's 2 X 3 between-participants factorial design with 3 repeated measures (vignettes), two-tailed $\alpha = .05$. Sign-ups were limited to people who had not completed any of our previous studies in this research, who held citizenship in the U.S. or Canada, who currently resided in one of those countries, and who had indicated in Prolific's prescreening survey that they had managerial experience. Participants were only permitted to begin the study if they were not on a mobile device and if they passed a reading-comprehension question. We received responses from 621 people. Following the pre-

registration, we dropped individuals who had IP addresses outside the U.S. or Canada, or who submitted data from duplicate IP address or Prolific IDs. If a participant failed a comprehension check about a vignette (described below), we dropped their responses to that particular vignette. We retained partial responses as planned, but could not analyze responses to a vignette if participants did not complete the dependent measure (disclosure). The final sample had 1,575 observations submitted by 562 unique participants (332 men, 230 women, 2 non-binary; $M_{\text{age}} = 35$ years, $SD = 9.94$).

Vignettes. Participants imagined themselves as the advisor in three vignettes, presented in randomized order (see Appendix, and verbatim materials online). In one, participants offer subordinates advice about whose submission they should choose as the winner of a product-naming contest; in another, they offer advice to a boss about a coworker's work performance; and in a third, they offer advice to peers on an ethics committee about whether the evidence suggested that a coworker had charged personal expenses to a corporate credit card. Thus, the vignettes varied the status of the advisee relative to participants. (The first two vignettes were the same as in Study 3, with minor wording changes).

Manipulations. Participants were randomly assigned to one of six between-participants conditions in a 2x3 design. They were in the same condition for all three vignettes. The *advice* manipulation varied whether participants offer the advisee a positive or a negative evaluation of the target person. For example, in the vignette about the product-naming contest, participants in the *positive-advice* condition read,

One of the ideas for a name is clearly the best in your opinion. You think it clever, creative and catchy. Thus, you advise the judges that you think this idea should be the winner.

whereas participants in the *negative advice* condition read,

One of the ideas for a name is clearly the worst in your opinion. You think it neither clever, nor creative, nor catchy. Thus, you advise the judges that you think this idea should not win.

The *feelings* manipulation varied whether the target person was someone participants liked, disliked, or did not have an opinion about. As in our prior studies, each vignette emphasized that the reason for these feelings had nothing to do with the performance participants had to evaluate. Thus, the feelings created an interpersonal conflict of interest. For example, in the vignette about the contest, participants read the following; brackets indicate text that varied by condition:

The judges don't know it, but this idea was submitted by an employee you know outside of work. You consider this person [a friend/an enemy/an acquaintance]. For reasons that have nothing to do with the quality of their ideas, you [find this person really agreeable/find this person really disagreeable/don't have an opinion about whether this person is agreeable or disagreeable].

Measures. For each vignette, participants completed the following measures. The comprehension check always came first, the dependent measure always came last, and the order of the remaining measures was randomized.

Comprehension check. Participants answered a simple multiple-choice question to check their understanding of the vignette (see Verbatim Online Materials).

Impression-management concerns: Sociability, competence, and morality. To measure expectations about how positively participants thought they would be perceived if they disclosed their feelings, we administered a validated, 9-item scale that captures the three fundamental dimensions on which people form impressions: sociability, competence, and morality (Leach et al., 2007). As in Study 3, the sociability subscale was our hypothesized mediator (H2a; items: likeable, friendly, and warm; $\alpha > .95$ in each vignette). The competence and morality subscales were exploratory measure (competence items: intelligent, competent, and skilled; $\alpha > .94$ in each

vignette; morality items: honest, sincere, and trustworthy; $\alpha > .94$ in each vignette). Participants responded to each item on a 7-point scale from 1 = *Not at all* to 7 = *Very much*. We randomized the order in which participants completed the three subscales.

Advice credibility. To measure participants' concern with appearing credible, we asked them to imagine that they had disclosed their feelings about the target person to their advisees. Then they rated what the advisees would think about their evaluation of the target's performance on four 7-point, bipolar items: accurate/inaccurate, untrustworthy/trustworthy, unreliable/reliable, and correct/incorrect ($\alpha > .92$ in each vignette). We reverse-coded the first and last items so that higher numbers would mean that participants expected their advice to seem more credible.

Concerns about others' objectivity. The same two items from Study 3 asked participants how disclosure would affect the relevant advisees' judgment. Specifically, participants indicated on 5-point scales how difficult it would be for the advisees to remain objective and the extent to which disclosure would affect the advisees' judgment ($r > .64$ in each vignette). We included this measure as a potential mediator (H2b). As a second test of this mediation prediction, we added a new four-item measure on which participants indicated whether disclose would make the advisees' judgment less/more fair, less/more just, less/more accurate, and less/more correct ($\alpha > .94$ in each vignette).

Morality of disclosure. As an exploratory measure, we asked participants to rate disclosure on four 7-point, bipolar items; immoral/moral, unethical/ethical, the wrong thing to do/the right thing to do ($\alpha > .94$ in each vignette). We report the results of this measure in the Online Supplement, and refer to them in Study 5's Discussion section.

Dependent measure: likelihood of disclosure. Finally, participants indicated how likely they would be to disclose, using the 5-point scale from our prior studies.

Results

We pre-registered one-tailed significance test for all directional predictions (Cho & Abe, 2013), but we report two-tailed tests for simplicity because they produced identical conclusions. Table 2 shows correlations among all measures.

Advisors are reluctant to disclose dislike. Replicating our central finding and consistent with H1, participants said they were less likely to disclose that they disliked the target person ($M = 2.96$, $SD = 1.22$) than that they liked the target person ($M = 3.31$, $SD = 1.15$), $d = .30$. This main effect of the feelings manipulation was significant in a multi-level regression model with fixed effects for the type of feelings ($-1 =$ disliking, $0 =$ neutral, $1 =$ liking), the type of advice ($-1 =$ negative, $1 =$ positive), their interaction, and vignettes (dummy-coded), plus random intercepts for participants, $b = .18$, $SE(b) = .06$, $z = 2.98$, $p = .003$ (see Table 3).⁵

⁵ The results were virtually identical when we reran the analysis without the neutral condition.

Table 2

Correlations Among Study 5's Measures

	1	2	3	4	5	6	7	8
1. sociability	1							
2. morality	0.62*	1						
3. competence	0.74*	0.72*	1					
4. credibility	0.39*	0.41*	0.44*	1				
5. influence	-0.03	-0.18*	-0.11*	-0.29*	1			
6. fair	0.31*	0.32*	0.35*	0.42*	-0.19*	1		
7. right_to_do	0.37*	0.45*	0.39*	0.17*	-0.05	0.28*	1	
8. disclosure	0.35*	0.41*	0.38*	0.19*	-0.05*	0.28*	0.61*	1

Note. Variables 1–3 measure different impression-management concerns. Variable 5 (how much disclosure would influence advisees' judgments) and 6 (whether disclosure would make advisees' judgments more or less fair) measure concerns about others' objectivity. Variable 7 measures whether participants think disclosure is the right thing to do. Variable 8 is the main dependent measure, disclosure intentions. * $p < .05$

Table 3

Regression Results for Study 5's Disclosure Measure

	<i>b</i>	<i>SE(b)</i>	<i>z</i>	<i>p</i>	[95% CI]	
Feelings	0.18	0.06	2.98	0.003	0.06	0.29
Advice	-0.05	0.05	-1.09	0.274	-0.15	0.04
Feelings X Advice	0.13	0.06	2.17	0.030	0.01	0.24
Vignette ethics	0.25	0.05	4.76	0.000	0.15	0.35
contest	-0.23	0.05	-4.39	0.000	-0.34	-0.13
Constant	3.23	0.06	56.93	0.000	3.12	3.34

Note: Coding for the feelings manipulation: -1 = disliking, 0 = neutral, 1 = liking. Coding for advice manipulation: -1 = negative, 1 = positive. "Ethics" and "contest" are dummy codes for vignette, with the "performance evaluation" vignette as the omitted reference group. Random effect parameters are omitted.

We next examined whether these results reflect a reluctance to disclose negative feelings, an eagerness to disclose positive feelings, or both. These analyses were pre-registered as exploratory because we did not have a hypothesis. The results only supported the idea that participants were reluctant to disclose negative feelings. Specifically, participants were less willing to disclose negative feelings than neutral feelings, ($M = 3.42$, $SD = 1.23$), $b = .49$, $SE(b) = .12$, $z = 4.11$, $p < .001$, $d = .39$, but they were not significantly more willing to disclose positive feelings than neutral feelings, $b = .14$, $SE(b) = .12$, $z = 1.25$, $p = .227$, $d = .10$. These results are from a mixed regression model with dummy codes for the feelings condition (coding the neutral condition as the reference group), an effect code for type of advice, fixed effects for vignettes, and random intercepts for participants.

Additional exploratory analyses suggested that participants were reluctant to disclose dislike to advisees, regardless of whether the advisees were subordinates, peers, or a boss. As noted, participants' relationship with the advisee varied by vignette. The same significant preference for concealing dislike (compared to liking or neutral feelings) emerged in all three vignettes (see Online Supplement).

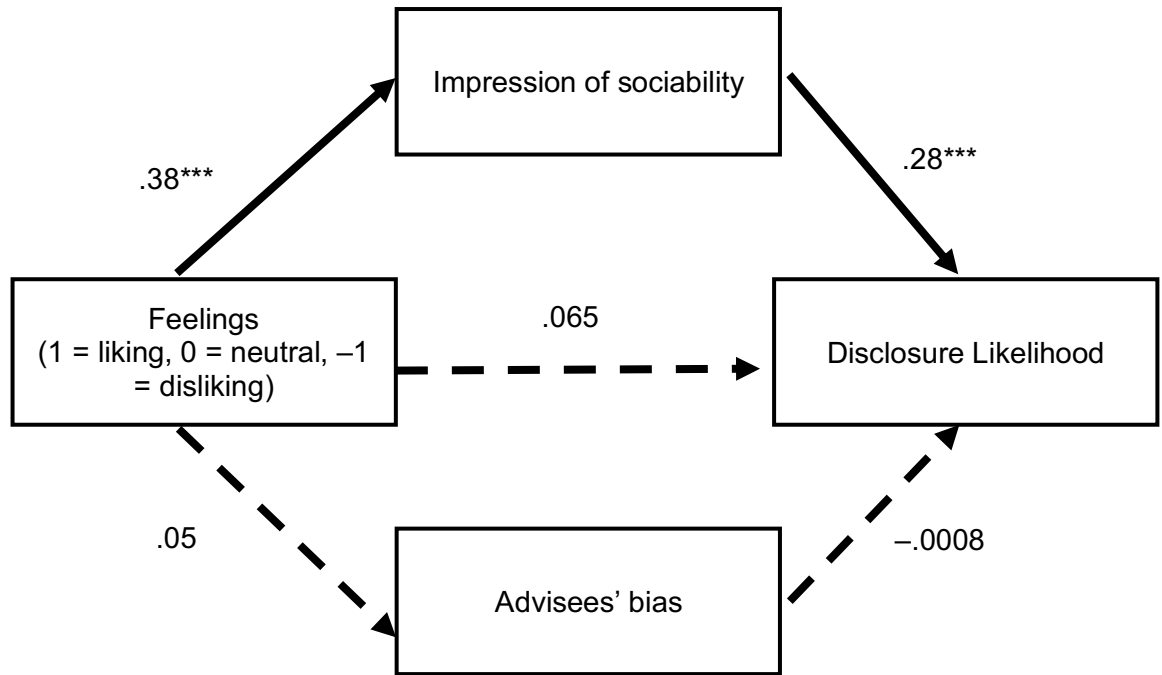
Impression-management concerns about sociability mediate reluctance to disclose dislike. Consistent with H2a, people's reluctance to disclose negative feelings was significantly mediated by impression-management concerns about sociability, $b = .11$, $SE(b) = .02$, $z = 5.36$, $p < .001$. In other words, people thought that disclosing dislike would make them appear less sociable than disclosing liking, which in turn predicted greater reluctance to disclose (see Figure 4's top path). Inconsistent with H2b, their reluctance was not significantly mediated by concerns about others' objectivity $b = -.00003$, $SE(b) = .001$, $z = .03$, $p = .979$. Participants did not expect that disclosing disliking would bias advisees' judgments any more than disclosing liking would

(see Figure 4's bottom path). These results, which directly replicate Study 3 and conceptually replicate Study S2 (see Online Supplement), emerged in a pre-registered, multi-level, parallel mediation analysis, with z -scored mediators, fixed effects for feelings ($-1 =$ disliking, $0 =$ neutral, $1 =$ liking), advice ($-1 =$ negative, $1 =$ positive), and vignettes (dummy-coded), and random intercepts for participants. This analysis used the two-item measure of concerns about others' objectivity from Study 3, but the conclusions were identical when we replaced it with the new four-item measure.⁶ Thus, the specific items used to measure concerns about objectivity in Study 3 cannot explain why this variable was not a significant mediator.

⁶ Specifically, the indirect effect through sociability remained significant, $b = .09$, $SE(b) = .02$, $z = 5.10$, $p < .001$, and the indirect effect through the new (z -scored) measure of objectivity concerns was not significant, $b = .006$, $SE(b) = .01$, $z = .60$, $p = .546$.

Figure 4

Impression-Management Concerns About Sociability Mediate the Reluctance to Disclose Dislike in Study 5



Note. Solid lines indicate statistically significant effects. Mediators were z-scored; disclosure likelihood was measured on a 5-point scale. *** $p < .001$

Additional analyses (not pre-registered) supported our claim that impression-management concerns specifically about sociability – above and beyond concerns about the other two fundamental dimensions of impressions, morality and competence – can explain why advisors were more reluctant to disclose negative than positive feelings that pose conflicts of interest. Specifically, when we added the measures of morality and competence (both *z*-scored) to the parallel mediation model just described, the indirect effect through sociability concerns remained significant, $b = .04$, $SE(b) = .014$, $z = 2.94$, $p = .003$.⁷

As noted, our theorizing does not make any predictions about whether the indirect effects through morality and competence would be significant. Exploratory analyses showed no significant effect through impression-management concerns about morality, $b = .014$, $SE(b) = .010$, $z = 1.41$, $p = .158$. We did find some evidence for an indirect effect through concerns about competence, $b = .016$, $SE(b) = .008$, $z = 1.97$, $p = .048$. That is, people thought they would appear more competent if they disclosed a positive relationship than a negative relationship – and the more competent they expected to appear, the more inclined they were to disclose. However, this unexpected effect should be interpreted with caution because it emerged in one of two exploratory analysis with a *p*-value just below .05, and is no longer significant after applying a Bonferroni correction (which lowers the critical *p*-value from .05 to .025).

Additional exploratory analyses showed that participants expected that they would appear less sociable if they disclosed dislike ($M = 3.93$, $SD = 1.20$) versus neutral feelings ($M = 4.72$, $SD = 1.16$), $z = 5.17$, $p < .001$, $d = .53$. They did not expect that they would appear more sociable if they disclosed liking ($M = 4.83$, $SD = 1.21$) compared to neutral feelings, $z = .85$, $p = .394$, $d = .08$. These analyses followed the same method as our analysis of the disclosure measure.

⁷ The indirect effect through concerns about others' objectivity remained non-significant, $b = .001$, $SE(b) = .002$, $z = .71$, $p = .480$.

In sum, the evidence suggests that impression-management concerns about appearing unsociable – above and beyond concerns about appearing moral, appearing competent, or about others' objectivity – can explain people's reluctance to disclose dislike.

Exploring the role of credibility concerns. We next explored two questions about advisors' desire to appear credible and its role in disclosure decisions (not pre-registered). First, did participants think that their advice would seem more credible if they disclosed feelings that conflicted with their evaluation of the target person's performance? Figure 5 suggests the answer is yes. The interaction shown between the feelings manipulation ($-1 = \text{negative}$, $0 = \text{neutral}$, $1 = \text{positive}$) and the advice manipulation advice ($-1 = \text{negative}$, $1 = \text{positive}$) was significant in a multi-level regression model with fixed effects for vignettes (dummy-coded) and random intercepts for participants, $b = -.60$, $SE(b) = .07$, $z = 8.64$, $p < .001$ (see Table 4). Participants thought positive evaluations of the target person's performance would seem more credible if they disclosed that they disliked ($M = 4.76$, $SD = 1.43$) versus liked ($M = 3.80$, $SD = 1.26$) or felt neutrally about the target person ($M = 4.26$, $SD = 1.29$), $z_s = 5.09$ and 2.72 , respectively, $ps < .001$. By contrast, they thought that a negative evaluation would seem more credible if they disclosed that they liked ($M = 4.74$, $SD = 1.43$) versus disliked ($M = 3.34$, $SD = 1.35$) or felt neutrally about the target person ($M = 4.69$, $SD = 1.29$), $z_s = 7.36$ and 6.99 , respectively, $ps < .001$.

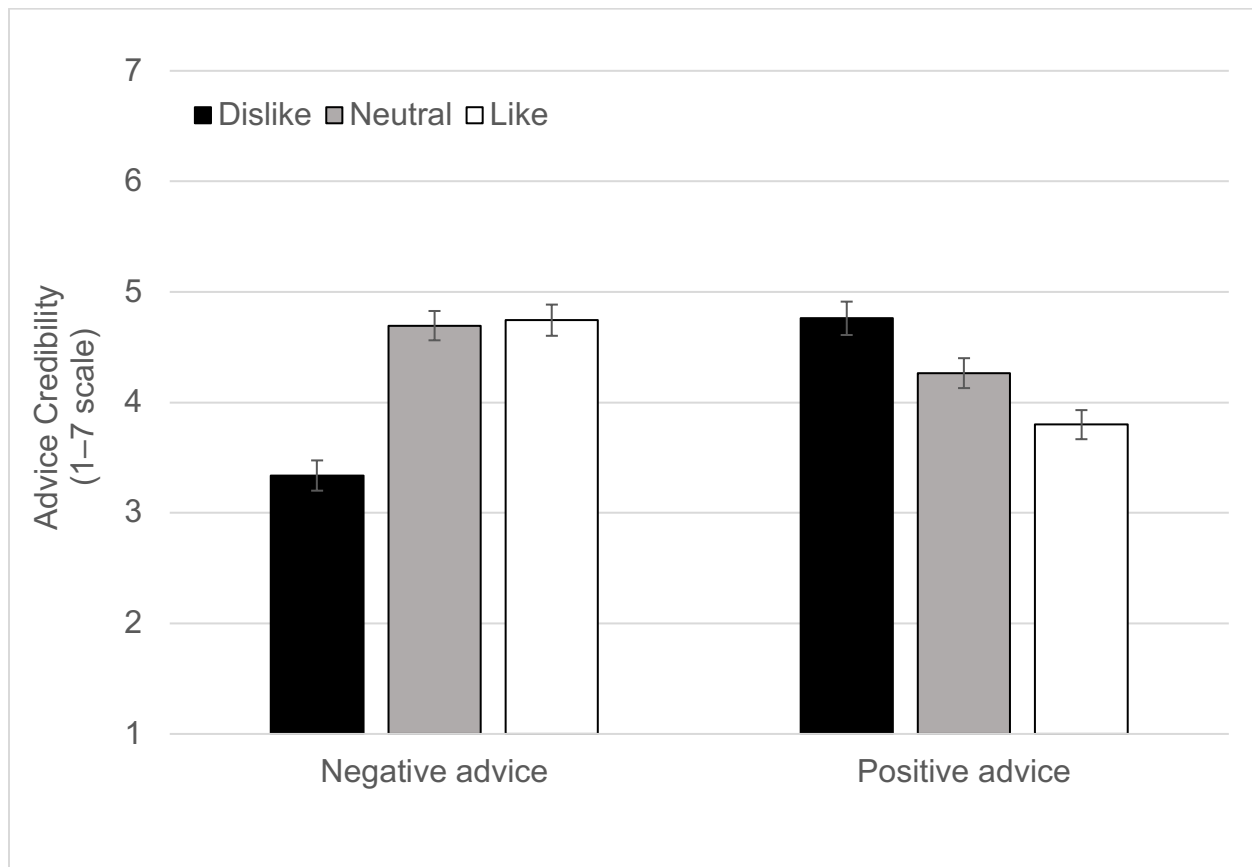
Figure 5Advice Credibility: Means \pm SEs in each of Study 5's Conditions

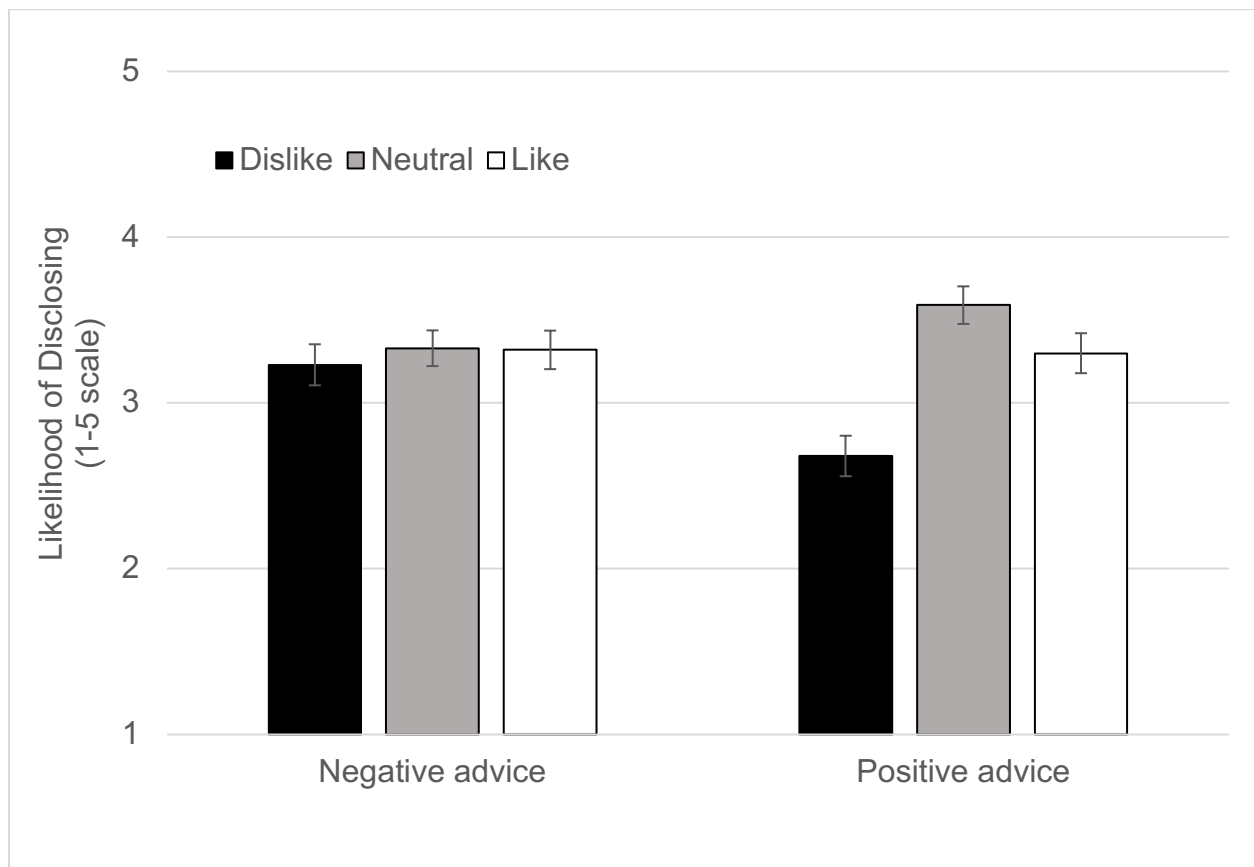
Table 4

Regression Results for Study 5's Credibility Measure

	<i>b</i>	<i>SE(b)</i>	<i>z</i>	<i>p</i>	95% CI for <i>b</i>	
	0.10	0.07	1.41	0.159	-0.04	0.23
Feelings	0.00	0.06	-0.01	0.995	-0.11	0.11
Advice						
Feelings X						
Advice	-0.60	0.07	-8.64	0.000	-0.73	-0.46
Vignette						
ethics	0.00	0.07	0.02	0.985	-0.14	0.14
ideas	-0.32	0.07	-4.38	0.000	-0.46	-0.17
Constant	4.35	0.07	62.62	0.000	4.21	4.49

Note: Coding for the feelings manipulation: -1 = disliking, 0 = neutral, 1 = liking. Coding for advice manipulation: -1 = negative, 1 = positive. "Ethics" and "contest" are dummy codes for vignette, with the "performance evaluation" vignette as the omitted reference group. Random effect parameters are omitted.

Second, did these concerns about credibility influence participants' disclosure decisions? If so, then according to the results just presented, participants should be more inclined to disclose interpersonal conflicts of interest when their feelings and evaluations are misaligned. For example, they should be more likely to disclose dislike in the positive-advice condition (misaligned) than in the negative-advice condition (aligned). However, as Figure 6 shows, we found no support for this possibility. In fact, participants were *less* likely to disclose dislike in the positive-advice condition (misaligned; $M = 2.68$, $SD = 1.16$) than in the negative-advice condition (aligned; $M = 3.23$, $SD = 1.22$), $z = 3.31$, $p < .001$ (compare Figure 6's black bars), and their willingness to disclose liking did not significantly depend on whether their advice was positive (aligned; $M = 3.30$, $SD = 1.15$) or negative (misaligned; $M = 3.32$, $SD = 1.15$), $z = .19$, $p = .85$. (see Figure 5's white bars). In other words, the interaction between the feelings and advice manipulation was significant, $b = .13$, $SE(b) = .06$, $z = 2.17$, $p = .030$ in the multi-level regression analysis of the disclosure measure described previously (see Table 3), but the shape of this interaction did not map onto participants' credibility concerns (refer back to Figure 5).

Figure 6Likelihood of Disclosure: Means \pm SEs in each of Study 5's Conditions

Probing this interaction further revealed that participants who imagined offering positive advice were more reluctant to disclose disliking ($M = 2.68$, $SD = 1.16$) than liking ($M = 3.30$, $SD = 1.15$) or neutral feelings ($M = 3.59$, $SD = 1.08$), $z = 3.63$, $p < .001$, $d = .54$, and $z = .28.76$, $p < .001$, $d = .82$, respectively (see Figure 6's right-hand bars) — but by contrast, participants who imagined offering negative advice were not significantly more reluctant to disclose dislike ($M = 3.23$, $SD = 1.22$) than liking ($M = 3.32$, $SD = 1.15$) or neutral feelings ($M = 3.59$, $SD = 1.08$), $z_s < .59$, $p_s > .5$, $d_s < .10$ (see Figure 6's left-hand bars). This finding raises the possibility that advisors' reluctance to disclose negative feelings about a target person does not generalize to situations in which they have offered an unambiguously negative evaluation of the target's performance. However, there is reason to question the robustness of this unpredicted pattern. The p -value of the relevant interaction effect ($p = .030$) underestimates the Type-I error rate because it was just one of several exploratory analyses (see Online Supplement and pre-registration document); adjusting for multiple comparisons would render the effect non-significant. Moreover, unlike the main effect of feelings manipulation, this interaction was not significant in all three vignettes (see Online Supplement). Despite these caveats, we speculate on why this pattern may have emerged in the Discussion section below.

Discussion

Study 5's results provide further evidence of our central finding, and address several remaining questions. First, the results suggest that advisors are reluctant to disclose that they dislike the target, but not eager to disclose that they like the target. Participants expected that disclosing whom they disliked would make them appear dislikeable, but did not expect that disclosing whom they liked could make them appear likeable. These expectations are reasonable, given that social norms proscribe speaking negatively of others (Bergsieker et al., 2012), and that

violating norms leaves a more negative impression more than complying with norms leaves a positive impression (Reeder & Spores, 1983; Risky & Birnbaum, 1974).

Second, the results generalized across three different relationships an advisor could have with an advisee: subordinate, peer, and boss.

Third, Study 5 shed additional light on mechanism with mediation analyses. Concerns about appearing unsociable were able to explain participants' reluctance to disclose dislike above and beyond concerns about appearing immoral or incompetent (H2a). Despite using a new measure of objectivity concerns, we again found no evidence for the hypothesized that such concerns would shape disclosure decisions (H2b). We also found no evidence that concerns about appearing credible drove participants' inclination to disclose. Participants were not more likely to disclose feelings that were misaligned (vs. aligned) with their advice, despite expecting that doing so would make their advice seem more credible. Some evidence hinted at the interesting possibility that concerns about appearing incompetent – in *addition* to concerns about appearing unsociable – could also help explain people's reluctance to disclose dislike. However, this evidence was not robust, and it does not challenge our explanation based on sociability concerns.

Unexpectedly, participants' reluctance to disclose dislike in Study 5 was less pronounced (and not significant) when they imagined giving their advisee a negative evaluation of the target person's performance. As noted, there are statistical reasons to question this result's robustness. Moreover, this result did not replicate in Study 2, in which evaluations of the target person's performance did not significantly moderate preferences for disclosing disliking (vs. liking). These caveats aside, one explanation is that Study 5's participants were mindful that a negative evaluation could harm the target person, and thus felt morally obligated to adopt a higher

standard of transparency when offering a negative evaluation. Consistent with this possibility, participants were more likely to think that disclosing dislike was the moral thing to do when they had evaluated the target negatively versus positively (see Online Supplement).

General Discussion

Advisors who must objectively evaluate the performance of someone they like or dislike face an interpersonal conflict of interest: Their feelings towards this person could cloud their judgment about the person's performance. Five studies (plus four more described in the Online Supplement) suggest that advisors are less likely to disclose disliking than liking or neutral feelings (H1).

These findings were robust across multiple populations and paradigms, including four pre-registered experiments. MBA students from a class with 64 different nationalities were less likely to say they had actually disclosed real interpersonal conflicts of interest at work when those conflicts arose because they disliked versus liked a coworker (Study 1), regardless of the strength of these feelings. The same students were also less inclined to disclose disliking than liking if those feelings created a conflict of interest in the future. Urban residents with 64 different nationalities were more likely to actually disclose experimentally induced liking (vs. disliking) that posed a conflict of interest in the laboratory (Study 2), again regardless of the strength of those feelings. Finally, online participants in the U.S., Canada, the U.K., and Ireland, as well as undergraduate business majors in the US and MBA students from a class with 66 different nationalities anticipated being less likely to disclose interpersonal conflicts of interest stemming from disliking (vs. liking) in a variety of workplace scenarios (Studies 3–5, and S1a-S1c). Evidence for this phenomenon emerged when the advisees to whom advisors could

disclose were peers, bosses, or subordinates (see especially Study 5). Together, these studies provide robust evidence for a novel phenomenon.

Mechanisms

Impression-management concerns about sociability. The studies also point to a reason why advisors are reluctant to disclose dislike: They worry that disclosure would make advisees perceive them as unsociable (H2a). This mechanism, which drew convergent support from three measurement-of-mediation studies (Studies 3, 5, and S2) and a moderation-of-process experiment (Study 4), builds on research showing that audiences tend to dislike those who express dislike of others (e.g., Amabile, 1983; Ames et al., 2010). Advisors thus appear to conceal dislike as a strategy to manage the impressions they leave on advisees.

Alternative explanations. Advisors could have multiple reasons for preferring to disclose positive instead of negative interpersonal conflicts of interest, but the current data did not support several other mechanisms.

Other impression-management concerns. Consistent with our theorizing, Study 5 found that impression-management concerns about sociability mediated advisors' reluctance to disclose disliking (vs. liking) above and beyond impression-management concerns about morality and competence, the other fundamental dimensions of person-perception (Brambilla & Leach, 2014; Goodwin, 2015). Moreover, Study 5's results were inconsistent with the possibility that advisors disclosed to enhance their credibility. This evidence supports our view that advisors' reluctance to disclose dislike arises from a concern about appearing unsociable in particular rather than looking bad in general.

Our theorizing and this evidence, however, do not preclude the possibility that other impression-management concerns could play a role as well. For example, Study 5 found some

evidence – albeit not statistically robust – that advisors worried they would appear less competent if they disclosed disliking than if they disclosed liking, and that these worries predicted disclosure decisions. Future research should examine whether this effect replicates. Study 5 found no evidence that impression-management concerns about morality could explain advisors' reluctance to disclose dislike, but future research might find support for other, unmeasured mediators.

Concerns about advisees' objectivity. We also found no support for a mechanism based on concerns about advisees' objectivity, failing to confirm H2b. On two different measures, advisors did not indicate that disclosing whom they disliked would bias advisees' judgments significantly more than disclosing whom they liked (Studies 3 and 5). We also did not find any support for H2b when we manipulated participants' concerns about the advisee's objectivity in a pre-registered supplemental study with five vignettes ($N = 798$; see Study S2 in the Online Supplement). In one vignette, for example, the advisee (a coworker) asks participants for their opinion about which job candidate to hire. Depending on randomly assigned, between-participant conditions, participants either liked or disliked the candidate, and the advisee either had influence or no influence over the hiring decision. We reasoned that participants should care less about the advisee's objectivity when the advisee had no influence. Thus, H2b predicts that participants' reluctance to disclose disliking, relative to liking, should be weaker when the advisee has no influence. However, the results showed no support for this prediction. Participants were just as reluctant to disclose that they disliked (vs. liked) the target person, regardless of how much they should worry about the advisee's objectivity. In short, multiple, independent tests of H2a found no evidence that concerns about others' objectivity explained the results.

Other alternative explanations. We also did not find evidence that advisors strategically conceal dislike to get away with giving the advisee a biased evaluation of the target person (Study 2), or that advisors base their disclosure decisions on what advisees are already likely to know (Studies 1 and 2; also Studies S1a–S1c, which emphasized that the advisee was unaware of the advisors’ feelings; see Appendix S1 in the Online Supplement). Results presented in the Online Supplement do not support the possibility that advisors believe it would be harder to remain objective if they disliked than if they liked someone (Studies S1a–S1c, and Study 3’s supplemental analyses).

Whereas the impression-management mechanism focuses on how people imagine disclosure will make them *appear*, another potential alternative explanation focuses on how people imagine disclosure will make them *feel*. This explanation, however, cannot account for Study 4’s finding that advisors’ were less reluctant to disclose disliking relative to liking when they imagined that they did not care what their advisees thought of them. Concerns about how one feels could be part of the story, but concerns about how one appears are necessary to fully explain the effect.

Potential Boundary Condition

Study 5 found evidence for an unexpected boundary condition: Participants were only reluctant to tell their advisee that they disliked (vs. liked) the target person when they imagined that they had evaluated the advisee’s performance negatively. As noted, one explanation is that participants felt morally obligated to disclose dislike when they evaluated the performance negatively, even though they also knew that such a disclose would make them appear less sociable and credible. Perhaps mindful that negative evaluations can cause harm, participants appear to have adopted a higher standard of transparency in this case.

However, several considerations mentioned in Study 5 raise questions about the robustness of this effect. Additionally, lab participants in Study 2 were more reluctant to disclose negative than positive feelings even when they gave more negative evaluations of the target person's performance. Thus, although the results make us most confident that people are reluctant to disclose dislike when they have offered a positive evaluation of the target person, more research is needed to understand whether this effect is really smaller or even absent when they have offered a negative evaluation.

Theoretical Contributions

Our results advance understanding of conflicts of interest and advice-giving in several ways. First, they begin to address a need to better understand how advisors decide to disclose their conflicts of interest. Prior work has focused on how conflicts of interest bias advisors' judgments (e.g., Moore et al., 2010) and on how advisees respond when they learn about such conflicts (e.g., Cain et al., 2011), but we know little about when advisors will tell advisees about these conflicts. Descriptively, we find that many people choose to disclose conflicts of interest. The majority of Study 1's participants said they had disclosed interpersonal conflicts of interest at work, and the majority of Study 2's participants actually disclosed such conflicts of interest in the lab. Our main contribution is to show that the size of the majority that chooses to disclose depends on whether the conflict involves liking versus disliking.

Second, our findings answer a recent call to consider interpersonal dynamics in advice-giving contexts (Blunden et al., 2019). Most work on advice examines when and to what extent advice improves advisees' accuracy, but neglects the role of impression management and other interpersonal motives (for reviews, see Bonaccio & Dalal, 2006; Van Swol et al., 2018). The work that does consider these motives focuses on why advisees seek and take advice

(MacGeorge, Feng, & Guntzviller, 2016; Sah et al., 2013), and why advisors offer advice (Blunden & Gino, 2018; Effron & Miller, 2015) and avoid conflicts of interest (Sah & Loewenstein, 2014). Our research reveals how interpersonal concerns affect advisors in two novel respects: Interpersonal conflicts of interest are less likely to come to light when they arise from disliking instead of liking, and advisors' motivation to manage advisees' impressions may help explain this effect.

Third, the results suggest that conflict-of-interest theories should be broadened to encompass a wider range of situations. Current theories aim to explain how advisors' desire to benefit themselves financially can compromise their objectivity and influence their advisees' decisions (e.g., Cain et al., 2005; Chugh et al., 2005; Moore et al., 2006; Sah, 2019). In many real-world conflicts of interest, however, the threat to advisors' objectivity instead comes from a desire to benefit someone they like or to harm someone they dislike. Although existing theories have little to say about these interpersonal conflicts of interest, such conflicts are ubiquitous. Many teams, organizations, and professions are small enough that members inevitably get asked to evaluate the performance of someone with whom they have a personal relationship. The present research illustrates how considering these interpersonal conflicts of interest can spark new discoveries that are not captured by existing theories, such as an asymmetry in preferences for disclosing liking versus disliking.

Practical Implications

Our studies suggest that if advisees want to know about interpersonal conflict of interest, they should actively seek information about whom an advisor might dislike. Such information is less likely to spontaneously come to light than information about whom an advisor likes. If an organization wishes to encourage disclosure of negative relationships that could pose a conflict

of interest, our theorizing suggests it should consider ways of minimizing advisors' impression-management concerns. Consider again the employee who dislikes a coworker for reasons unrelated to his work performance. The employee would probably be more likely to disclose this negative relationship if she had to evaluate the coworker's performance on an anonymous form than if she had to do it in a public meeting. Fostering psychological safety could also allow advisors feel comfortable disclosing whom they like without worrying as much about appearing dislikable themselves (see Edmondson, 1999).

Limitations and Future Directions

By shifting the literature's focus from financial conflicts of interest to interpersonal conflicts of interest, the present work suggests several questions for future research. First, do judgments and behavior in conflict-of-interest situations depend on whether a financial incentive versus an interpersonal feeling is responsible for the conflict? For example, future research could investigate which type of conflict of interest (financial vs. interpersonal) advisors are more likely to disclose, which one people expect to exert more bias on judgment, and which one actually exerts more bias. The answer to such questions has important implications for understanding when and why conflicts of interest will go undetected in organizations.

Second, how do *advisees* think about advisors' interpersonal conflicts of interest? It will be important to understand how advisees expect such conflicts to influence advisors, how much advisees are willing to tolerate such conflicts, when advisees want advisors to disclose, when advisees predict advisors will disclose, and how advisees respond to such disclosure. For example, a central question in the literature is whether disclosing an advisor's conflict of interest makes advisees more or less likely to follow the advice (e.g., Cain et al., 2011; Crawford & Sobel, 1982; Sah, 2019). The answer to this question may differ if the conflict of interest is

interpersonal rather than financial. Advisors might feel more pressure to follow advice if they knew it would help their advisor's friend than if they knew it would help their advisor's pocketbook.

Third, previous research suggests that advisors' impression-management concerns about disclosing whom they dislike are warranted (e.g., Ames et al., 2010), but are these concerns well-calibrated? Future research should investigate whether advisors overestimate how negatively advisees would react to such disclosures (see John et al., 2016).

Fourth, does advisors' reluctance to disclose whom they dislike depend on how much they dislike the person? As noted, even mild liking or disliking can bias how we evaluate people's performance (e.g., Landy & Sigall, 1974), just like small financial incentives can skew our judgments (Dana, 2009). Studies 2 and 3 found no evidence that the strength of advisors' feelings about the target person moderated their preferences for disclosing liking versus disliking. However, future research should investigate this issue more systematically by manipulating the intensity of advisors' liking and disliking. Advisors might not consider extremely weak liking or disliking to constitute a conflict of interest, and thus see no reason to disclose either.

Fifth, is advisors' greater reluctance to disclose negative (vs. positive) interpersonal conflicts of interest bad for advisees' decision-making? The answer to this question is unclear, even though advisees want advisors to disclose both types of conflict. To answer the question, we would need to know whether advisors offer more-biased advice about people they dislike than about people they like, and whether advisees discount the advisors' advice differently depending on whether the advisor has a negative versus positive relationship with the target. Future research should explore these issues.

A final question concerns the generalizability of our results. The present studies examined conflicts of interest in real, simulated, and hypothetical advice-giving situations that people could realistically encounter in a work context (e.g., completing a performance evaluation of a coworker). Our international participant samples in Studies 2, 3, and 5 suggest some generalizability across cultural contexts (over 60 different nationalities were represented), although a systematic examination of cross-cultural differences must await future research. Future research is also necessary to test whether the results generalize to professional advisors, such as lawyers, brokers, or doctors. Finally, we did not assess how advisors expected their advisees to feel about the relevant target person. Knowing that the advisee likes or dislikes the target person would complicate advisors' decisions to disclose their own liking or disliking for the target person. Future research should explore such situations.

Conclusion

Advisors often have discretion about whether to disclose interpersonal feelings that could compromise their judgment. Our research suggests that advisors are more inclined to disclose some feelings than others. When interpersonal conflicts of interest arise, advisors are more likely to reveal whom they like than whom they dislike.

References

- Amabile, T. M. (1983). Brilliant but cruel: Perceptions of negative evaluators. *Journal of Experimental Social Psychology, 19*(2), 146-156.
- Ames, D. R., Bianchi, E. C., & Magee, J. C. (2010). Professed impressions: What people say about others affects onlookers' perceptions of speakers' power and warmth. *Journal of Experimental Social Psychology, 46*(1), 152-158.
- Arkin, R. M. (1981). Self-presentation styles. In J. T. Tedeschi (Ed.), *Impression management theory and social psychological research* (pp. 311-334). New York: Academic Press.
- Armor, D. A. (1999). The illusion of objectivity: A bias in the perception of freedom from bias. *Dissertation Abstracts International, 59*(9-B).
- Ashford, S. J., & Northcraft, G. B. (1992). Conveying more (or less) than we realize: The role of impression-management in feedback-seeking. *Organizational Behavior and Human Decision Processes, 53*(3), 310-334.
- Babcock, L., & Loewenstein, G. (1997). Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic Perspectives, 11*(1), 109-126.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323-370.
- Bazerman, M. H., & Tenbrunsel, A. E. (2012). *Blind spots: Why we fail to do what's right and what to do about it*. Princeton, NJ: Princeton University Press.
- Bergsieker, H. B., Leslie, L. M., Constantine, V. S., & Fiske, S. T. (2012). Stereotyping by omission: Eliminate the negative, accentuate the positive. *Journal of Personality and Social Psychology, 102*(6), 1214.

- Blunden, H., & Gino, F. (2018). How the other half thinks: The psychology of advising. In E. L. MacGeorge & L. M. Van Swol (Eds.), *The Oxford Handbook of Advice* (pp. 43-68). New York: Oxford University Press.
- Blunden, H., Logg, J. M., Brooks, A. W., John, L. K., & Gino, F. (2019). Seeker beware: The interpersonal costs of ignoring advice. *Organizational Behavior and Human Decision Processes, 150*, 83-100.
- Bolino, M., Long, D., & Turnley, W. (2016). Impression management in organizations: Critical questions, answers, and areas for future research. *Annual Review of Organizational Psychology and Organizational Behavior, 3*, 377-406.
- Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes, 101*(2), 127-151. doi: 10.1016/j.obhdp.2006.07.001
- Brambilla, M., & Leach, C. W. (2014). On the importance of being moral: The distinctive role of morality in social judgment. *Social Cognition, 32*(4), 397-408.
- Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues, 55*(3), 429-444. doi: 10.1111/0022-4537.00126
- Brooks, A. W., Gino, F., & Schweitzer, M. E. (2015). Smart people ask for (my) advice: Seeking advice boosts perceptions of competence. *Management Science, 61*(6), 1421-1435.
- Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (Don't expect an easy answer). *Journal of Personality and Social Psychology, 98*(4), 550-558.
- Buttelmann, D., & Böhm, R. (2014). The ontogeny of the motivation that underlies in-group bias. *Psychological Science, 25*(4), 921-927.

- Cain, D. M., Loewenstein, G., & Moore, D. A. (2005). The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *Journal of Legal Studies*, *34*, 1-25.
- Cain, D. M., Loewenstein, G., & Moore, D. A. (2011). When sunlight fails to disinfect: Understanding the perverse effects of disclosing conflicts of interest. *Journal of Consumer Research*, *37*(5), 836-857.
- Cardy, R. L., & Dobbins, G. H. (1986). Affect and appraisal accuracy: Liking as an integral dimension in evaluating performance. *Journal of Applied Psychology*, *71*(4), 672.
- Carlston, D. E., & Skowronski, J. J. (2005). Linking versus thinking: Evidence for the different associative and attributional bases of spontaneous trait transference and spontaneous trait inference. *Journal of Personality and Social Psychology*, *89*(6), 884.
- Cho, H.-C., & Abe, S. (2013). Is two-tailed testing for directional research hypotheses tests legitimate? *Journal of Business Research*, *66*(9), 1261-1266. doi: <https://doi.org/10.1016/j.jbusres.2012.02.023>
- Chugh, D., Bazerman, M. H., & Banaji, M. R. (2005). Bounded ethicality as a psychological barrier to recognizing conflicts of interest. In D. A. Moore, D. M. Cain, G. F. Loewenstein & M. H. Bazerman (Eds.), *Conflicts of interest: Challenges and solutions in business, law, medicine, and public policy* (pp. 74-95). New York, NY: Cambridge University Press.
- Chugh, D., & Kern, M. C. (2016). A dynamic and cyclical model of bounded ethicality. *Research in Organizational Behavior*, *36*, 85-100.
- Church, B. K., & Kuang, X. (2009). Conflicts of interest, disclosure, and (costly) sanctions: Experimental evidence. *The Journal of Legal Studies*, *38*(2), 505-532.

- Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica: Journal of the Economic Society*, 50(6), 1431-1451. doi: 10.2307/1913390
- Dana, J. (2009). How psychological research can inform policies for dealing with conflicts of interest in medicine. In B. Lo & M. J. Field (Eds.), *Conflict of interest in medical research, education, and practice*. Washington, DC: National Academies Press.
- Dana, J., & Loewenstein, G. (2003). A social science perspective on gifts to physicians from industry. *JAMA*, 290(2), 252-255. doi: 10.1001/jama.290.2.252
- De Dreu, C. K., Greer, L. L., Handgraaf, M. J., Shalvi, S., Van Kleef, G. A., Baas, M., . . . Feith, S. W. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science*, 328(5984), 1408-1411.
- Dobbins, G. H., & Russell, J. M. (1986). The biasing effects of subordinate likeableness on leaders' responses to poor performers: A laboratory and a field study. *Personnel Psychology*, 39(4), 759-777.
- Dubrin, A. J. (2011). *Impression management in the workplace*. New York: Routledge.
- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350-383.
- Effron, D. A., & Miller, D. T. (2015). Do as I say, not as I've done: Suffering for a misdeed reduces the hypocrisy of advising others against it. *Organizational Behavior and Human Decision Processes*, 131, 16-32. doi: 10.1016/j.obhdp.2015.07.004
- Elsevier. (2019, March). FACTSHEET: Competing Interests. Retrieved April 1, 2020, from https://www.elsevier.com/_data/assets/pdf_file/0007/653884/Competing-Interests-factsheet-March-2019.pdf

Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal.

Journal of Applied Psychology, 66(2), 127.

Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of*

Experimental Social Psychology, 47(6), 1231-1236.

Fiske, S. T., Cuddy, A. J., 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.

Folkes, V. S., & Sears, D. O. (1977). Does everybody like a liker? *Journal of Experimental*

Social Psychology, 13(6), 505-519.

Gardner, W. L., & Martinko, M. J. (1988). Impression management in organizations. *Journal of*

Management, 14(2), 321-338.

Gawronski, B., & Walther, E. (2008). The TAR effect: When the ones who dislike become the

ones who are disliked. *Personality and Social Psychology Bulletin, 34*(9), 1276-1289.

Goodwin, G. P. (2015). Moral character in person perception. *Current Directions in*

Psychological Science, 24(1), 38-44.

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-

168. doi: 10.1037/a0034726

Guo, T., Sriram, S., & Manchanda, P. (2019). 'Let the sun shine in': The impact of industry

payment disclosure on physician prescription behavior. Retrieved from

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2953399

- Halevy, N., Bornstein, G., & Sagiv, L. (2008). "In-group love" and "out-group hate" as motives for individual participation in intergroup conflict: A new game paradigm. *Psychological Science, 19*(4), 405-411.
- Halevy, N., Weisel, O., & Bornstein, G. (2012). "In-group love" and "out-group hate" in repeated interaction between groups. *Journal of Behavioral Decision Making, 25*(2), 188-195.
- Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes, 70*(2), 117-133. doi: 10.1006/obhd.1997.2697
- Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting Economics, 31*(1-3), 405-440. doi: 10.1016/S0165-4101(01)00018-0
- Hunt, J. S., & Budesheim, T. L. (2004). How jurors use and misuse character evidence. *Journal of Applied Psychology, 89*(2), 347-361.
- Hwong, A. R., Sah, S., & Lehmann, L. S. (2017). The effects of public disclosure of industry payments to physicians on patient trust: a randomized experiment. *Journal of General Internal Medicine, 32*(11), 1186-1192.
- 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. doi: 10.1073/pnas.1516868113
- Johnson, D. E., Erez, A., Kiker, D. S., & Motowidlo, S. J. (2002). Liking and attributions of motives as mediators of the relationships between individuals' reputations, helpful behaviors and raters' reward decisions. *Journal of Applied Psychology, 87*(4), 808.

- Jones, E. E., & Pittman, T. S. (1982). Toward a general theory of strategic self-presentation. In J. Suls (Ed.), *Psychological perspectives on the self* (Vol. 1, pp. 231-262). Hillsdale, NJ: Lawrence Erlbaum.
- Koch, C., & Schmidt, C. (2010). Disclosing conflicts of interest – Do experience and reputation matter? *Accounting, Organizations and Society*, 35(1), 95-107. doi: <https://doi.org/10.1016/j.aos.2009.05.001>
- Landy, D., & Sigall, H. (1974). Beauty is talent: task evaluation as a function of the performer's physical attractiveness. *Journal of Personality and Social Psychology*, 29(3), 299.
- Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology*, 93(2), 234-249. doi: 10.1037/0022-3514.93.2.234
- Leary, M. R. (1995). *Self-presentation: Impression management and interpersonal behavior*. Oxford, UK: Westview Press.
- MacGeorge, E. L., Feng, B., & Guntzviller, L. M. (2016). Advice: Expanding the communication paradigm. *Annals of the International Communication Association*, 40, 239-270. doi: 10.1080/23808985.2015.11735261
- Mae, L., Carlston, D. E., & Skowronski, J. J. (1999). Spontaneous trait transference to familiar communications: Is a little knowledge a dangerous thing? *Journal of Personality and Social Psychology*, 77(2), 233.
- Maj, M. (2008). Non-financial conflicts of interests in psychiatric research and practice. *British Journal of Psychiatry*, 193(2), 91-92. doi: 10.1192/bjp.bp.108.049361
- McPherson Frantz, C. (2006). I AM being fair: The bias blind spot as a stumbling block to seeing both sides. *Basic and Applied Social Psychology*, 28(2), 157-167.

- Messick, D. M., & Sentis, K. P. (1979). Fairness and preference. *Journal of Experimental Social Psychology, 15*(4), 418-434.
- Moore, D. A., & Loewenstein, G. (2004). Self-interest, automaticity, and the psychology of conflict of interest. *Social Justice Research, 17*(2), 189-202.
- Moore, D. A., Tanlu, L., & Bazerman, M. H. (2010). Conflict of interest and the intrusion of bias. *Judgment and Decision Making, 5*(1), 37.
- Moore, D. A., Tetlock, P. E., Tanlu, L., & Bazerman, M. H. (2006). Conflicts of interest and the case of auditor independence: Moral seduction and strategic issue cycling. *Academy of Management Review, 31*(1), 10-29.
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153-163.
- Pham-Kanter, G., Alexander, G. C., & Nair, K. (2012). Effect of physician payment disclosure laws on prescribing. *Archives of Internal Medicine, 172*(10), 819-821.
- PLoS Medicine Editors. (2008). Making sense of non-financial competing interests. *PLoS Medicine, 5*(9), e199. doi: 10.1371/journal.pmed.0050199
- Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends in Cognitive Sciences, 11*(1), 37-43.
- Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review, 111*(3), 781-799.
- Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality & Social Psychology Bulletin, 28*(3), 369-381.

- Reeder, G. D., & Spores, J. M. (1983). The attribution of morality. *Journal of Personality and Social Psychology*, *44*(4), 736-745.
- Regan, D. T., Straus, E., & Fazio, R. (1974). Liking and the attribution process. *Journal of Experimental Social Psychology*, *10*(4), 385-397.
- Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, *7*(4), 331. doi: 10.1037/1089-2680.7.4.331
- Riskey, D. R., & Birnbaum, M. H. (1974). Compensatory effects in moral judgment: Two rights don't make up for a wrong. *Journal of Experimental Psychology*, *103*, 171-173. doi: 10.1037/h0036892
- Rose, S. L., Sah, S., Dweik, R., Schmidt, C., Mercer, M., Mitchum, A., . . . Robertson, C. (2019). Patient responses to physician disclosures of industry conflicts of interest: A randomized field experiment. *Organizational Behavior and Human Decision Processes*.
- Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296-320.
- Sah, S. (2019). Conflict of interest disclosure as a reminder of professional norms: Clients first! *Organizational Behavior and Human Decision Processes*, *154*, 62-79.
- Sah, S., Fagerlin, A., & Ubel, P. (2016). Effect of physician disclosure of specialty bias on patient trust and treatment choice. *Proceedings of the National Academy of Sciences*, *113*(27), 7465-7469.
- Sah, S., & Loewenstein, G. (2014). Nothing to declare: Mandatory and voluntary disclosure leads advisors to avoid conflicts of interest. *Psychological Science*, *25*(2), 575-584. doi: 10.1177/0956797613511824

- Sah, S., Loewenstein, G., & Cain, D. M. (2013). The burden of disclosure: Increased compliance with distrusted advice. *Journal of Personality and Social Psychology, 104*(2), 289-304.
doi: 10.1037/a0030527
- Saver, R. S. (2012). Is it really all about the money? Reconsidering non-financial interests in medical research. *The Journal of Law, Medicine & Ethics, 40*(3), 467-481.
- Schwartz, J., Luce, M. F., & Ariely, D. (2011). Are consumers too trusting? The effects of relationships with expert advisers. *Journal of Marketing Research, 48*(SPL), S163-S174.
- Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin, 105*(1), 131-142.
- Skowronski, J. J., Carlston, D. E., Mae, L., & Crawford, M. T. (1998). Spontaneous trait transference: Communicators take on the qualities they describe in others. *Journal of Personality and Social Psychology, 74*(4), 837.
- Smith, G. (2012, March 14). Why I am leaving Goldman Sachs, *The New York Times*. Retrieved from <https://www.nytimes.com/2012/03/14/opinion/why-i-am-leaving-goldman-sachs.html>
- Smith, R. (2006). Conflicts of interest: How money clouds objectivity. *Journal of the Royal Society of Medicine, 99*(292-297).
- Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology, 89*(6), 845.
- Tenbrunsel, A. E., & Messick, D. M. (2004). Ethical fading: The role of self-deception in unethical behavior. *Social Justice Research, 17*(2), 223-236.

- Thompson, D. F. (1993). Understanding financial conflicts of interest. *New England Journal of Medicine*, 329(8), 573-576. doi: 10.1056/NEJM199308193290812
- Turban, D. B., Jones, A. P., & Rozelle, R. M. (1990). Influences of supervisor liking of a subordinate and the reward context on the treatment and evaluation of that subordinate. *Motivation and Emotion*, 14(3), 215-233.
- Van Swol, L. M., Paik, J. E., & Prahll, A. (2018). Advice recipients: The psychology of advice utilization. In E. L. MacGeorge & L. M. Van Swol (Eds.), *The Oxford Handbook of Advice* (pp. 21-42). New York: Oxford University Press.
- Wang, C. S., Galinsky, A. D., & Murningham, J. K. (2009). Bad drives psychological reactions, but good propels behavior. *Psychological Science*, 20, 634-644.
- Wayne, S. J., & Ferris, G. R. (1990). Influence tactics, affect, and exchange quality in supervisor-subordinate interactions: A laboratory experiment and field study. *Journal of Applied Psychology*, 75(5), 487.
- Weisel, O., & Böhm, R. (2015). "Ingroup love" and "outgroup hate" in intergroup conflict between natural groups. *Journal of Experimental Social Psychology*, 60, 110-120.
- Westfall, J. (2016). *PANGEA: Power ANalysis for GEneral Anova designs*. Retrieved from <http://jakewestfall.org/publications/pangea.pdf>
- Whitson, J. A., Wang, C. S., See, Y. H. M., Baker, W. E., & Murnighan, J. K. (2015). How, when, and why recipients and observers reward good deeds and punish bad deeds. *Organizational Behavior and Human Decision Processes*, 128, 84-95.
- Wyer, R. S., Budesheim, T. L., & Lambert, A. J. (1990). Cognitive representation of conversations about persons. *Journal of Personality and Social Psychology*, 58(2), 218.

Appendix

Vignettes in Studies 3–5

Note. The text in brackets below was manipulated. Due to differing time constraints, Study 3 administered the first two vignettes shown below, Study 4 administered just the first one, and Study 5 administered all three. To provide evidence of generalizability, Studies S1a–S2 administered these three plus two more (see Online Supplement, Appendix S1 and S2). Study 5 made minor wording changes (e.g., to specify that the advisee was a boss, peer, or subordinate; see verbatim materials at https://osf.io/npzw4/?view_only=e1fe082d5f54479fb3074492daf5da82).

Contest

Suppose you work for a company that is holding a contest to name its new product. Any employee can submit an idea for a name. The employee whose idea is chosen will win \$5,000. The company has narrowed down the ideas to a small pool of finalists. You have been randomly selected as one of eight judges who will vote on the winning idea from this pool.

It just so happens that one of the ideas was submitted by someone with whom you have a **[positive/negative]** relationship outside of work. That is, you have a strongly **[positive/negative]** relationship with this person, for reasons unrelated to the quality of their idea.

Performance Evaluation

Suppose one of your colleagues is up for a big promotion at work. To help determine whether this person is qualified for the promotion, your boss asks you and several other employees whether you would each like to fill out an evaluation of this person's work performance.

The boss doesn't know it, but this person happens to be someone with whom you have a **[positive/negative]** relationship with outside of work. That is, you have a strongly **[positive/negative]** relationship with this person, for reasons unrelated to their work performance.

[Note: Study 5 revised the vignette so that the colleague was up for a bonus rather than a promotion]

Ethics

Suppose you have been randomly selected to be on your company's twelve-person ethics committee. The committee's job is to review alleged wrongdoing by employees and determine

their guilt or innocence. The committee has just been asked to review the case of an employee who is accused of charging personal expenses to the company credit card.

By coincidence, this person is an **old [friend/enemy]** from outside of work. For reasons completely unrelated to their guilt or innocence, you find this person extremely **[likeable/dislikeable]**.