You Rate Me and I'll Rate You:
Mutual Rating Relationships in Multi-Rater Performance Evaluation Systems

Martin Artz
martin.artz@wiwi.uni-muenster.de
School of Business and Economics
University of Münster

Carolyn Deller
cdeller@wharton.upenn.edu
The Wharton School
University of Pennsylvania

Sinja Leonelli
sinja.leonelli@stern.nyu.edu
Stern School of Business
New York University

May 2025

Abstract

We examine rating relationship formation and rating behavior in a multi-rater performance evaluation system. Specifically, we study mutual rating relationships, where two employees rate each other contemporaneously. We use proprietary data from an online retailer and show that demographic similarity, organizational proximity, and strong cooperation incentives are positively associated with the likelihood of a mutual rating relationship. Mutual ratings are higher on average than one-sided ratings and this premium is more pronounced when there is relationship persistence and when employees have strong personal incentives to cooperate for higher ratings. Supervisors place more weight on mutual vis-à-vis one-sided ratings when there are relatively fewer ratings available for the employee being evaluated. This higher perceived informativeness of mutual ratings aligns with our finding that supervisors are less likely to veto employee rater nominations that would lead to a mutual rating relationship. Overall, our study offers first evidence on a phenomenon inherent to multi-rater systems: mutual rating relationships.

We thank Jen Choi (discussant), Jungho Choi, Anna Costello, Fei Du (discussant), Wayne Guay, Jonas Heese, Mirko Heinle, Christopher Ittner, Roope Keloharju, Michal Matějka (discussant), Jochen Pierk, and Evelien Reusen (discussant) for their comments and suggestions, as well as seminar participants at the 2021 Erasmus Accounting Workshop, the 2022 Hawaii Accounting Research Conference, the 2022 Management Accounting Section Midyear Meeting, the 2022 GMARS Conference, the 2023 INSEAD Accounting Symposium, the TRR 266 Seminar Series at Paderborn University, the University of Pennsylvania, the National University of Singapore, the University of Utah, Stanford University, the University of Chicago, Harvard Business School, and Tilburg University. We sincerely thank the research site for access to data and helpful discussions. Carolyn Deller gratefully acknowledges the financial support received as a Dorinda and Mark Winkelman Distinguished Faculty Scholar. All errors are our own.

I. INTRODUCTION

Many firms use multi-rater performance evaluation systems, whereby an employee is evaluated by multiple individuals (e.g., peers, direct reports, senior managers) in addition to their main supervisor. The idea is that such systems provide a more comprehensive view of performance and result in greater validity relative to a single supervisor assessment, particularly in modern workplaces characterized by decentralization, autonomy, and remote work (London and Smither 1995; Conway, Lombardo, and Sanders 2001; Di Fiore and Souza 2021). Nonetheless, supervisors may still have considerable influence on performance evaluations in such systems. For example, supervisors may determine who provides an evaluation or be responsible for arriving at an overall rating, ideally informed by the multi-rater assessments and their own view (WorldatWork 2019).

In this study, we examine an unexplored phenomenon inherent to multi-rater systems used in firms: *mutual rating relationships*. We define a mutual rating relationship (hereafter, "MRR") as a rating relationship in which two employees rate each other contemporaneously in the multi-rater system. We refer to *non-mutual* rating relationships as *one-sided* rating relationships. MRRs are likely to be ubiquitous in multi-rater systems and ratings from MRRs (hereafter "mutual ratings") may differ systematically from ratings from one-sided rating relationships (hereafter "one-sided ratings"). Understanding any such systematic differences is important as performance ratings are often used as the basis for compensation and personnel (e.g., promotion) decisions, and ratings can impact employees' motivation and fairness perceptions (Bol 2008; Capelli and Conyon 2008). Using proprietary company data, we examine (1) the formation and determinants of MRRs, (2) whether there are systematic differences between mutual and one-sided ratings, and (3) how mutual and one-sided ratings are weighted by supervisors in their overall evaluations.

Our study utilizes data from an online platform retailer that uses a multi-rater system—

with pay and career consequences—for thousands of white-collar employees each year. While employees nominate raters, raters are ultimately chosen by supervisors, who have the discretion to not only decide which employee nominations result in a rating request but to also add their own raters. If they wish, employees can decline a rating request (for instance, if they receive numerous requests). The company does not formally encourage nor restrict MRRs. Employees know who rated them, but individual ratings are not disclosed. Following the multi-rater rating stage, supervisors arrive at an overall rating, subject to review by a calibration committee (Demere, Sedatole, and Woods 2019). Such systems are common in practice.[1]

We first consider MRR formation. We expect a potential rater is more likely to be nominated in the presence of (1) demographic similarity (as there should be, in theory, greater interpersonal attraction and liking, making a favorable evaluation more likely), (2) organizational proximity (as it should enhance observability of performance and thus improve evaluation accuracy), or (3) strong cooperation incentives for compensation and promotion purposes (as this could enable mutually beneficial uprating). Moreover, and given the focus of our study, we hypothesize that these factors increase the likelihood of a nomination being reciprocated—that is, a *mutual* nomination.

Nominations do not solely determine MRRs, however, as supervisors and designated raters play a role in rating relationship formation. For our study, we have data corresponding to a period at our research setting when supervisors could veto nominated raters and also add their own raters. Whether supervisors impact the prevalence of MRRs through their formal actions is ex ante

[1] A recent survey (WorldatWork 2019) finds 48% of companies use raters other than the employee's direct supervisor in their formal performance evaluation process. Employees are often involved in the selection of their raters—42% of companies indicated that raters are selected by "Supervisor/Employee," "Supervisor with Suggestions," or "Employee." In 57% of the companies surveyed, employees know the identities of their raters. Supervisor rating discretion is common in these systems; in 50% of the companies surveyed, supervisors have complete flexibility as to how they incorporate feedback from other raters into their overall performance rating.

unclear, creating tension in our study. On the one hand, supervisors may use their vetoing power and rater additions to constrain MRRs if they believe that MRRs will result in less accurate evaluations, for instance, due to reciprocal uprating. On the other hand, supervisors may enable MRRs if they believe that mutual nominations arise between employees who know each other and their respective roles best and thus may yield the most accurate evaluations. It is also possible that supervisor vetoing of proposed MRRs could be perceived as a signal of mistrust, making supervisors reluctant to engage in such vetoing if they fear that employee commitment to the system (such as participation by providing ratings) might suffer as a result. In the final stage of rating relationship formation, we expect that designated raters will be less likely to decline a rating request that can lead to a MRR. This prediction rests on the assumption that mutual rating requests generally reflect employees' preferred rating relationships (notwithstanding supervisor involvement in rater selection).

It is an open question as to whether mutual ratings differ systematically from one-sided ratings, that is, if mutual ratings are higher or lower on average. The non-random nature of mutual vis-à-vis one-sided rater nominations and rating relationships could lead to systematic differences in ratings. For instance, mutual ratings could be higher than one-sided ratings due to greater interpersonal attraction and liking between employees in a MRR (provided supervisors have not vetoed such relationships). Mutual ratings could also be higher due to strategic rating behavior, in that there may be an implicit or explicit agreement between raters in a MRR to intentionally inflate each other's ratings (reciprocal uprating). However, there are several reasons why mutual ratings may be equal to, or even lower than, one-sided ratings on average.

First, if MRRs predominantly arise between employees who work more closely together and thus are more familiar with each other and the relevant performance expectations, mutual

ratings may be the most accurate ratings and there is no reason to expect such accuracy to translate to higher ratings. Second, strategic rating behavior may not occur if employees fear reputational costs, expect to be penalized if supervisors suspect less-than-truthful reporting (Grabner, Künneke, and Moers 2020), or have an aversion to lying and feel obligated to provide a true assessment (Evans, Hannan, Krishnan, and Moser 2001; Abeler, Nosenzo, and Raymond 2019). These two arguments suggest no systematic differences between mutual and one-sided ratings. Third, if employees who work more closely together perceive themselves to be in competition, and these employees tend to be in mutual rating relationships, this could lead to a downward bias in mutual ratings and thus higher one-sided ratings (Huang, Shum, Wu, and Xiao 2019). Finally, prior research shows that lack of familiarity can lead to more lenient (i.e., higher) ratings (Bol 2011); if one-sided ratings are indicative of less familiarity, one-sided ratings could be higher on average.

Finally, how supervisors weight mutual and one-sided ratings when arriving at their overall rating is an empirical question. Supervisors may weight mutual and one-sided ratings differently if they perceive that systematic differences between these ratings exist and they wish to factor such differences into their own rating. Alternatively, supervisors may place equal weight on mutual and one-sided ratings if no systematic differences exist, if they are cognizant that systematic differences exist but choose not to override them, or if distinguishing between these ratings requires time and effort that supervisors are unwilling to invest.

We use detailed data from the retailer's multi-rater system (for one or two periods, depending on the analysis) combined with demographic data to study these issues. Demographic similarity, organizational proximity, and strong cooperation incentives increase the likelihood of a potential rater being nominated—and within an employee's nomination set, demographic similarity and organizational proximity increase the likelihood of a *mutual* nomination.

When we investigate the nomination vetoing stage, we find that supervisors are *less likely* to veto a rater nomination that would result in a mutual (versus a one-sided) rating relationship. This vetoing behavior suggests that, despite any selection effects and the potential for strategic rating behavior, supervisors may nonetheless expect mutual ratings to be more accurate than one-sided ratings. Another potential explanation is that vetoing mutual nominations could lead to negative employee reactions (assuming mutual nominations reflect employees most preferred raters), which supervisors wish to avoid. Indeed, in a follow-up analysis, we find that employees are more likely to decline rating requests when their mutual nominations have been vetoed by a supervisor, that is, they are less willing to participate in the multi-rater system. Examining the raters that supervisors added, we find the likelihood that an addition results in a mutual (as opposed to a one-sided) rating request is higher in the presence of organizational proximity between employees, indicating that supervisors are more likely to themselves form mutual rating relationships when employees likely possess an informational advantage. In the last stage of the rating relationship formation process, we find, as expected, that potential raters are less likely to decline rating requests that would result in MRRs.

Mutual ratings are higher on average than one-sided ratings. Since "true" performance is unobservable (as is often the case when performance is evaluated subjectively), we do not know whether mutual or one-sided ratings better reflect an employee's underlying performance. However, our evidence is consistent with at least part of the mutual rating premium (i.e., mutual ratings exceeding one-sided ratings on average) being driven by strategic rating behavior under certain conditions. Specifically, we find the premium is more pronounced when employees arguably expect the MRR to persist in the future, and when both of the employees in the MRR have strong personal incentives to cooperate for compensation and promotion purposes.

Lastly, we find some evidence that supervisors, on average, place greater weight on mutual than one-sided ratings when forming their overall evaluations of employees. Further analyses reveal strong evidence of this differential weighting in the subsample of employees where supervisors have relatively fewer ratings for the employee under consideration, but no evidence of such differential weighting in the subsample where supervisors have relatively more ratings for the employee. This pattern suggests that, under limited information, one-sided ratings may be perceived as noisier, prompting supervisors to anchor on mutual ratings that reflect more reciprocal relationships and signal greater interpersonal familiarity.

Our research contributes to the literature on performance evaluation systems—specifically, multi-rater systems. Prior work primarily focuses on multi-rater systems used for developmental purposes; such systems have a long history and are often referred to as "360-degree feedback systems" (e.g., Deller, Gallani, and Sandino 2024). We examine a contemporary multi-rater system with pay and career consequences, and we depart from prior studies with our specific focus on MRRs. We identify conditions under which MRRs are more likely to form (i.e., when there is demographic similarity, organizational proximity or strong cooperation incentives), extending broader research on peer selection (e.g., Kiessling, Radbruch, and Schaube 2019; Kiessling, Radbruch, and Schaube 2022). We are the first to provide evidence of a mutual rating premium, and we document that strategic rating inflation drives at least part of this premium under certain conditions. Prior research, primarily in non-corporate settings, has shown that personal incentives and rating motives affect rating behavior but does not differentiate between mutual and one-sided ratings (e.g., Wong and Kwong 2007; Wang, Wong, and Kwong 2010; Huang et al. 2019; Riedl,

Grad, and Lettl 2024).[2] Our findings are based on real workplace performance evaluation data and show that the magnitude of the mutual rating premium depends on relationship persistence and the strength of cooperation incentives. In this vein, we contribute to the literature documenting strategic behavior by employees in response to different management control system designs (e.g., Towry 2003; Luft 2016; Hecht, Maas, and van Rinsum 2023).

We contribute to the literature on supervisor discretion in performance evaluation. Prior research has examined how supervisors weight different types of performance measures (e.g., financial versus non-financial measures), supervisor biases in performance evaluation, and heterogeneity in supervisor rating behavior (e.g., Ittner, Larcker, and Meyer 2003; Bol 2011; Du, Erkens, Young, and Tang 2018; Frederiksen, Kahn, and Lange 2020; Künneke 2024). Our study extends this literature in two ways. First, we show that supervisors use their discretion not only in evaluating performance but also in actively shaping multi-rater rating relationships—specifically, enabling or constraining MRRs. Second, we find that, at least under certain conditions, supervisors differentiate between ratings based on the nature of the underlying relationship (i.e., MRR versus one-sided) when making evaluation decisions. This evidence contributes to the emerging literature on supervisor behavior in multi-rater systems (e.g., Bol, Margolin, and Schaupp 2023) and broadens our understanding of discretion in performance evaluation beyond the traditional focus on performance measure properties such as financial versus non-financial or objective versus subjective indicators.

II. RELEVANT LITERATURE, HYPOTHESES, AND RESEARCH QUESTIONS

[2] Klapper, Piezunka, and Dahlander (2024) study an online community (Wikipedia) where evaluations can be made at any point in time, and thus, mutual versus one-sided rating relationships are not specified ex-ante as in our setting. While they cannot make the explicit distinction between rating relationships as we do, they find that raters with incentives to provide more evaluations focus their negative evaluations on those where the likelihood of negative reciprocity is low (i.e., where the relationship is more likely to be one-sided). In contrast, we identify circumstances where strategic rating behavior gives rise to a more pronounced mutual rating premium as both employees have an incentive to provide a more positive evaluation.

We begin by considering MRR formation. Next, we theorize as to why mutual ratings may differ systematically from one-sided ratings. Finally, we contend that supervisors may weight mutual ratings and one-sided ratings differently when arriving at their overall rating.

Mutual Rating Relationship Formation

A natural starting point is to identify circumstances under which a rating relationship is more likely to be mutual than one-sided. In a multi-rater system, the relationship formation process may begin with each employee nominating potential raters. We posit that mutual nominations are more likely in the presence of (1) demographic similarity, (2) organizational proximity, or (3) strong cooperation incentives.

In theory, employees are likely to nominate raters who they anticipate will give them (1) a relatively more favorable evaluation, or (2) a relatively more accurate evaluation (Brutus, Aucoin, and Petosa 2005). In so doing, employees are likely to nominate those with positive affect towards them (as this should translate to a more favorable evaluation) and those who are more familiar with them (as this should yield greater accuracy).[3] We use demographic similarity as one condition giving rise to the former, and organizational proximity as one condition giving rise to the latter.

Sociologists, social psychologists, and organizational scholars have long studied homophily, "the tendency of individuals to associate with similar others" (Lawrence and Shah 2020, 2). Preferences for homophily along demographic attributes have been documented in various settings, including the workplace (e.g., Marsen 1988; Shrum, Cheek, and Hunter 1988; Ibarra 1992). Pfeffer (1985, 69) explains this preference for demographic similarity as follows:

"… similarity is one of the most important bases of interpersonal attraction; and

[3] Brutus et al. (2005) find that, when asked to nominate raters in a *hypothetical scenario*, employees prefer to be rated by those who they perceive (1) have greater positive affect toward them and (2) are more familiar with them. However, the affect result manifests only for subordinates of the ratee and not peers. We build on this initial evidence by using archival field data to examine actual rater nominations from a firm's multi-rater system and by examining not only rater nominations in general, but *mutual* rater nominations specifically.

> demographic features such as age, race, and sex both help to determine similarity and also signal that those who share these features are more likely to be similar. People who share experiences and attitudes are more likely to like each other because they will understand each other better, and because liking someone who is similar is self-reinforcing …"

We expect demographic similarity to increase the likelihood of a potential rater being nominated. If employees wish to receive relatively more favorable evaluations, they should be more inclined to nominate individuals where mutual liking and interpersonal attraction are more pronounced (Antonioni and Park 2001; Sol 2016). In turn, demographically similar nominees should be more likely to return the nomination, that is, demographic similarity should also increase the likelihood of a rater nomination being reciprocated, i.e., a *mutual* (as opposed to a one-sided) rater nomination.[4,5] Employees might limit such nomination behavior, however, if they wish to avoid supervisor attention or scrutiny.

Employees who work more closely together should have a better understanding of each other's role and performance expectations and a greater opportunity to observe each other's work. This "organizational proximity" should lead to more accurate performance evaluations due to these informational advantages (Norton 1992; Brutus et al. 2005).[6] Thus, we expect organizational proximity to increase the likelihood of a potential rater being nominated, and the likelihood of a mutual rater nomination. A counterargument, however, is that competition (e.g., for bonuses) is likely to be more pronounced in the presence of organizational proximity. This could create incentives for employees to engage in sabotage (i.e., downgrading each other's ratings), and thus

[4] Reciprocity in nominations could arise due to an explicit agreement between employees who are demographically similar or arise naturally in the nomination process.

[5] Mutual rater nominations may also be more likely to arise between demographically similar employees to the extent that personality traits and behaviors vary with demographic characteristics. For instance, prior research finds that men (women) tend to be more agentic (communal) and that men tend to participate more in group discussions (Badura et al. 2018). Employees may prefer to be rated by those who act similarly to themselves.

[6] Some earlier studies examine the concept of "task acquaintance" in a typical supervisor-employee rating scenario. For instance, Kingstrom and Mainstone (1985) capture task acquaintance by having supervisors report their familiarity with specific dimensions of a sales representative's job. While "task acquaintance" is very similar in spirit to "organizational proximity" we choose the latter as our construct of interest as it better maps to our underlying variable.

reduce the willingness of employees to nominate raters who work closely with them.

Finally, we expect that strong cooperation incentives will influence rater nominations. We hypothesize that the likelihood of a mutual rater nomination is greater when both employees have strong incentives to engage in a MRR that enables reciprocal rating inflation. This is especially likely when both of the employees will benefit from an inflated rating in the current period and they do not directly compete for rewards from high(er) ratings like promotions or bonuses. In this case, employees have strategic reasons to nominate those that they can reasonably expect to reciprocate with favorable evaluations, creating a mutually reinforcing arrangement.

Overall, we expect demographic similarity, organizational proximity, and strong cooperation incentives to increase the likelihood of a mutual rater nomination, which we capture formally in our first three hypotheses:

H1a: Demographic similarity increases the likelihood of a mutual rater nomination.

H1b: Organizational proximity increases the likelihood of a mutual rater nomination.

H1c: Strong cooperation incentives increase the likelihood of a mutual rater nomination.

In many multi-rater systems, rating relationships are determined by both employees and supervisors. In such an environment, supervisors may be able to veto nominations and add additional raters. Whether supervisors will embrace or constrain MRRs is an empirical question. If supervisors are concerned that mutual rater nominations reflect relationships where interpersonal attraction and liking are most pronounced, and thus may lead to overly favorable evaluations, or if they are concerned that MRRs could give rise to strategic rating behavior, they may seek to constrain MRRs. Conversely, if supervisors believe that mutual rater nominations arise where employees know each other and the relevant job requirements best, they may expect ratings stemming from MRRs to be more informative. It is also possible that supervisor vetoing of

proposed MRRs could be perceived as a signal of mistrust, making supervisors reluctant to engage in such vetoing if they fear that employee commitment to the system (such as participation by providing ratings) might suffer as a result. Thus, we state our first research question as:

RQ1: Do supervisors impact the likelihood of mutual rating relationship formation?

An employee's decision to accept (or not) a rating request represents the final step in rating relationship formation. When employees have such discretion, we predict they are less likely to decline a request that is expected to result in an MRR—that is, one in which the employee will also be rated by the ratee. This prediction is based on the assumption that MRRs reflect employees' preferred rating relationships. Formally:

H2: A rating request is less likely to be declined by a designated rater when accepting is expected to result in a mutual rating relationship.

Mutual Rating Relationship Outcomes

We contend that mutual ratings may differ systematically from one-sided ratings; that is, they may be higher or lower on average. There are two non-mutually exclusive channels that could give rise to a mutual rating premium, i.e., mutual ratings being *higher* than one-sided ratings: (1) selection effects in relationship formation and (2) strategic rating behavior. Under (1), strategic rating behavior is not necessary—raters may give an accurate rating of the employee (from their perspective) that is higher than a one-sided rating due to a *selection effect*, that is, the non-random nature in which mutual rater nominations and MRRs arise. For instance, interpersonal attraction and liking can increase the likelihood of MRR formation (as opposed to a one-sided rating relationship) due to reciprocal rater nominations *and* increase the likelihood of rating each other more favorably (Kingstrom and Mainstone 1985; Tsui and Barry 1986). Prior research shows that similarity in demographics—which we expect to be one determinant of mutual rater nominations—is associated with greater interpersonal attraction and liking, frequency of

communication, and higher performance ratings (e.g., Speckbacher and Wiernsperger 2022; Tsui and O'Reilly 1989; Zenger and Lawrence 1989). Similarly, mutual rater nominations may be more likely when employees have had an especially favorable experience working together. Again, this should translate into mutual ratings being higher than one-sided ratings due to a selection effect.

In contrast to channel (1), under (2), there is an implicit or explicit agreement between raters in a MRR to intentionally inflate each other's ratings. Higher ratings may lead to favorable compensation or career consequences, creating an incentive to cooperate in quid-pro-quo inflation. Further, if employees expect that supervisors (or calibration committees) anticipate such inflation and will discount mutual ratings as a result, this can create an incentive for *all* employees in MRRs to engage in inflation, like the earnings management equilibrium outlined by Stein (1989).

Despite the arguments for a mutual rating premium stemming from selection effects, there are several reasons why there may be no such premium, or even a mutual rating discount. First, supervisors may veto mutual rater nominations where they suspect the proposed rater is likely to provide an overly favorable assessment (e.g., due to interpersonal attraction and liking). Second, prior research finds that higher information-gathering costs—akin to not knowing the ratee well—positively affects leniency bias, resulting in inflated ratings (Saal and Landy 1977; Bol 2011). To the extent that one-sided rating relationships reflect relationships where employees know each other less well, then any such leniency bias could lead to one-sided ratings being higher than mutual ratings. Third, if MRRs are most likely to form when employees know each other and the relevant job requirements well, then mutual ratings may arguably be more accurate and there is no reason to expect this to manifest as higher ratings. Employees who know each other and the job requirements best may even be in competition with each other, which could create an incentive to downgrade the rating given and cause mutual ratings to be lower than one-sided ratings.

As to the possibility of strategic rating behavior in MRRs, supervisors may veto mutual rater nominations where they expect such behavior to arise. Moreover, rating inflation can bear a psychological cost of misreporting; some individuals have strong preferences for honesty and exhibit an aversion to lying (Evans et al. 2001; Abeler et al. 2019). Strategic rating inflation could also lead to reputational costs and punishment. In this vein, Grabner et al. (2020) find that supervisors who inflate their employees' ratings receive lower ratings themselves. If employees suspect penalties from supervisors, this may constrain their willingness to purposely inflate ratings.

Overall, it is ex-ante ambiguous if and how mutual ratings will differ from one-sided ratings. We therefore pose the following research question:

RQ2: Do mutual ratings differ systematically from one-sided ratings?

To the extent that employees engage in strategic rating behavior in MRRs, economic theory suggests that the likelihood of such behavior increases with repeated rating interactions and favorable compensation or career consequences for the employees.

Evidence on repeated interactions suggests that the incentive for a rater to inflate a mutual rating increases with the belief that the other party will keep their promise. Specifically, both parties will cooperate (rather than individually reneging on their cooperation promise) if they believe the MRR will hold for an indeterminant number of periods and that the future gains of cooperation (i.e., receiving inflated ratings and/or *not* receiving lower ratings in the future as punishment for defection in the current period) outweigh any reputational costs or penalties, or the personal cost of violating one's honesty preferences (Dal Bó and Fréchette 2018; Raub, Buskens, and Frey 2019).[7] If employees in a MRR expect the relationship to persist in the future, the mutual

[7] We assume that employees can infer to some degree whether a rater has upheld their promise to "uprate" (i.e., inflate the rating given). As we explain in Section III, employees at our research site know who rated them (or should be able to readily obtain this information) and receive detailed feedback about their performance from their supervisor in the

gains of cooperation may exceed any penalties or costs, making strategic rating behavior more likely.[8] Similarly, we expect personal incentives, such as compensation and promotion incentives, to affect mutual rating behavior. Specifically, if both employees in a MRR will benefit from higher ratings, any incentive to engage in strategic rating behavior should be even more pronounced.

Our third research question is:

RQ3: Is there evidence of strategic rating behavior in mutual rating relationships?

Supervisor Ratings and Mutual Rating Relationships

In multi-rater systems, supervisors may be responsible for arriving at an employee's overall performance rating. Prior research emphasizes that supervisors' personal incentives and preferences can influence their ratings (e.g., Prendergast and Topel 1996; Prendergast 2002; Du, Tang, and Young 2012). While researchers have documented various supervisor biases (e.g., Lipe and Salterio 2000; Bol 2011), the literature also contains numerous examples of supervisor behavior in workplace settings that seem consistent, at least on average, with the assumed objectives of the firm. For example, Campbell (2008) shows that supervisors incorporate not only financial but also non-financial dimensions of store manager performance in promotion decisions. Deller (2023) documents that supervisors in a multinational company evaluate and promote employees in accordance with the principles of the company. Demere et al. (2019) find that calibration committees leave supervisor ratings unchanged 75% of the time, which they interpret as committee satisfaction with supervisor ratings.

If supervisors perceive systematic differences between mutual and one-sided ratings, they

final stage of the evaluation process. Moreover, there was no company policy prohibiting employees from discussing their ratings while the evaluation system was open for rating submission. Given these potential information sources, employees are likely able to deduce the ratings received from individual raters with a reasonable degree of confidence. The empirical evidence in Section IV provides support for this assumption (see the results described in footnote 23).

[8] In many situations a firm using a multi-rater system is likely unable to fully disentangle real performance from misreporting; thus, any expected penalties are unlikely to be substantial for an employee (Milgrom and Roberts 1992).

may place differential weight on these ratings when arriving at their overall evaluations. Prior research shows that supervisors use their discretion in weighting different performance measures when evaluating employees and determining career outcomes (Ittner et al. 2003; Campbell 2008; Grabner and Moers 2013). Due to selection effects and strategic rating behavior, mutual ratings may include not only a premium (a level effect) but also contain more noise compared to one-sided ratings.[9] If this is the case, supervisors may rely less on mutual ratings when arriving at their evaluation, though they could (instead) veto rater nominations that would result in a MRR.

Conversely, supervisors may put equal or more weight on mutual ratings if they view mutual ratings as equally or more informative than one-sided ratings, for instance, if mutual ratings primarily arise between employees who know each other best and therefore reflect greater interpersonal familiarity. It is also possible that supervisors weight mutual and one-sided ratings equally because they do not invest time and effort in identifying MRRs and therefore do not distinguish between mutual and one-sided ratings, or because there *are* systematic differences between these ratings but supervisors do not wish to override such differences.

We pose our final research questions as follows:

RQ4: Do supervisors (implicitly) weight mutual and one-sided ratings differently?

III. RESEARCH SITE AND DATA

Our research site is a European online platform retailer of fashion and lifestyle products.[10] The company has more than 10,000 employees and annual revenues of several billion euros. The company's multi-rater performance evaluation system was introduced in recent years to broaden the scope for employee ratings and applies to approximately 5,000 white-collar employees

[9] When we refer to the possibility of "selection effects" contributing to a mutual rating premium we mean any relationship-specific factors (e.g., unconscious favoritism or a favorable shared working experience) that result in a mutual rating being higher than a one-sided rating (absent strategic rating behavior).

[10] This study was deemed not to meet the criteria for engagement in human subjects research by one of our institutions.

working at company headquarters. Top management believed the system would stimulate an environment of learning, with the idea that feedback should be obtained from those who employees interact with in their daily work. The system was also intended to address fairness concerns and increase transparency. Before the multi-rater system, employees were evaluated only by their supervisor, and bonuses resulted from individually negotiated objectives between employee and supervisor. Our data pertains to the first two years that the multi-rater system was in use.

Performance Evaluation Process

The performance evaluation process has four stages. In the first stage, the company invites employees to nominate up to 15 potential raters (nominating more is possible) towards the end of the evaluation period.[11] In the first year, supervisors were instructed to select five of the nominated raters (employees were not provided with a formal list of raters until after the evaluation process, though they could have asked nominated raters if they were selected in the interim). In the second year, supervisors formally approved or vetoed each nominated rater, with employees receiving formal notification of any vetoes. Supervisors are instructed to consider exposure (i.e., whether an individual can provide accurate/representative feedback) and diversity (i.e., raters from different levels, departments, etc.) when determining the final list of raters. Supervisors can add raters (thus, rating requests comprise approved nominations and supervisor additions), and employees asked to provide a rating can decline. Company policy is that each employee should provide and receive approximately five ratings, but this is not strictly enforced (in the first year, most employees received five ratings; in the second year, we observe more variation). Raters evaluate performance on multiple dimensions (which may vary somewhat from year to year), some independent of and

[11] Individuals can also nominate themselves to rate another employee (though this rarely occurs).

some specific to the job type, and provide an overall rating, from 1 (lowest) to 5 (highest).[12] We use the overall rating in our analyses as it is comparable across jobs and years.

In the second stage, supervisors rate each employee as a 1 ("Needs Improvement"), 2 ("Good Performer"), or 3 ("Top Performer"). Supervisors see their employees' first stage ratings and are supported by an electronic system to exchange information with other supervisors. Supervisors have full discretion as to how they incorporate this information as well as their own assessment when arriving at an employee's rating. Supervisors are instructed to rate approximately 10% of their employees as a 1, 80% as a 2, and 10% as a 3, but they can deviate if necessary.[13]

In the third stage, calibration committees review the ratings for all employees, focusing on employees who receive a rating of 1 or 3 from their supervisor, those for whom the supervisor rating differs substantially from what would be expected given the first stage ratings (in such cases, the supervisor may have to personally justify their rating), and those considered ready for promotion. The committee determines the final rating via confirmation or adjustment of the supervisor rating. Employees with a final rating of 2 or 3 receive a salary increase of between 3% and 8% and meet the performance criterion to be considered "ready for promotion" (generally a necessary but insufficient condition for promotion). Employees with a final rating of 3 also qualify for a bonus equivalent to roughly two months of salary. Employees with a final rating of 1 receive neither a raise nor a bonus and cannot be rated "ready for promotion."[14]

In the fourth stage, the results of the evaluation process are discussed in a meeting between

[12] In the second year of the system, raters rated an employee's "promotion readiness" in addition to their performance, and employees provided a self-evaluation. Raters could also explain their ratings with comments in the second year.
[13] Such forced distribution systems are common in practice. While any such effects are outside the scope of our study, research highlights that these systems can have beneficial effects, such as increasing performance (Berger, Harbring, and Sliwka 2013), but also entail costs, such as higher stress for employees (Cardinaels and Feichter 2021). A series of cases on the company Henkel traces the evolution of such a system over time (Simons and Kindred 2012; Simons and Deller 2022a; Simons and Deller 2022b).
[14] The (European) regulatory environment of the company does not permit forced layoffs due to low performance evaluations. Thus, a final rating of 1 does not lead to forced turnover.

each employee and their supervisor. Employees may be informed of their average ratings from the first stage but individual ratings and comments from the first stage are not officially disclosed.

Data and Variable Measurement

The company shared two periods of anonymized data from the multi-rater system and demographic information for employees. The period 2 data is more comprehensive, capturing not only the first-stage and supervisor ratings, which the period 1 data is limited to, but also rater nominations, supervisor vetoes of nominations and their rater additions, as well as the status of rating requests (e.g., finished, declined). The data covers all employees who participated in the system, except for the highest hierarchical levels. We refer to the employee who was rated as the "receiver," the employee who provided a first-stage rating as the "giver," and the receiver's main supervisor as the "primary supervisor."[15] Below we describe the measurement of our main variables. Appendix A contains the full list of variable definitions.

Demographic Similarity. Following the literature on homophily (e.g., Pfeffer 1985), we use nationality, gender, age, and company tenure to capture demographic similarity. We construct an aggregate *Demographic Similarity* variable which equals 1 if the receiver and giver (potential/designated or actual) share at least three demographic characteristics (i.e., the majority); otherwise, 0.

Organizational Proximity. Organizational proximity for a receiver-giver pair is based on whether they share a supervisor, belong to the same department, or have the same hierarchical

[15] All employees have a primary supervisor, and depending on their department and job, an employee may have up to two secondary supervisors (e.g., management accounting employees may have a primary business unit supervisor and a secondary supervisor from centralized accounting). Based on discussions with the company, we replace the receiver's primary supervisor with the secondary supervisor responsible for personnel management (and treat that supervisor as the primary supervisor throughout) when the primary supervisor's employee ID is missing in the period 2 data (we cannot ascertain in period 1 which secondary supervisor is responsible for personnel management). At the receiver level, we make a total of 234 replacements (less than 5% of the receivers in period 2 with a non-missing ID number) in the raw data (i.e., before any sample selection takes place).

rank. We construct an aggregate *Organizational Proximity* variable which equals 1 if the receiver and giver have at least two in common (i.e., the majority); otherwise, 0.

Noncompetitive Promotion Prospects. We consider a receiver-giver pair to have strong cooperation incentives for compensation and promotion purposes if both are considered "ready for promotion" but they are not in direct competition with each other. Specifically, the variable *Noncompetitive Promotion Prospects* equals 1 if the receiver and giver are both considered "ready for promotion" but are in different departments and are at different ranks; 0 otherwise.

Rating Relationship Formation. *Nominee* equals 1 if a potential giver is nominated to rate the receiver; otherwise, 0. *Mutual Nomination* equals 1 if *Nominee* equals 1 *and* the receiver is nominated to rate the potential giver; equals 0 if *Nominee* equals 1 but the receiver is *not* nominated to rate the potential giver (i.e., it is a one-sided nomination). *Supervisor Vetoed Rater Nomination* equals 1 if a supervisor vetoed the nomination for the potential giver to rate the receiver; 0 if a supervisor approved the nomination. *Supervisor-Created Mutual Rating Request* equals 1 if a supervisor's rater addition resulted in a mutual rating request (i.e., the designated giver and the receiver both subsequently received a request to rate the other) and equals 0 if a supervisor's rater addition resulted in a one-sided rating request (i.e., the designated giver received a request to rate the receiver, but the receiver did not receive a request to rate the giver). *Rating Request Declined* equals 1 if the designated giver declined the request to rate the receiver and equals 0 otherwise. *R_Pct. Mutual Approved Nominations* is the percentage of the approved (by a supervisor) nominations for proposed givers to rate the receiver where there is a corresponding approved nomination for the receiver to rate the giver. *Mutual Approved Rating Request* equals 1 if the rating request is a mutual request; 0 if the rating request is a one-sided request.

Mutual Rating Relationships. *Mutual Rating Relationship* equals 1 if the giver rates the

receiver *and* the receiver rates the giver; 0 if the giver rates the receiver but the receiver *does not* rate the giver (i.e., it is a one-sided rating relationship). *Proposed MRR Vetoed* equals 1 if the giver rates the receiver but the nomination for the receiver to rate the giver was vetoed by a supervisor (i.e., there would have been a MRR if not for a veto); 0 otherwise. *MRR Both Periods* equals 1 if a MRR exists between the receiver and the giver in period 1 and period 2; 0 otherwise. *MRR Period 1 Only* (*MRR Period 2 Only*) equals 1 if a MRR exists between the receiver and the giver in period 1 but not in period 2 (in period 2 but not period 1); otherwise, 0.

Ratings. *Rating Received* is the overall rating (on a 1-5 scale) the giver gave the receiver in the first stage. *Supervisor Rating* is equal to 1 if the supervisor rated the receiver as "Needs Improvement"; 2 if "Good Performer"; and 3 if "Top Performer." *R_Weighted Mean Mutual Rating* (*R_Weighted Mean One-sided Rating*) is the average of the receiver's mutual (one-sided) ratings in the first stage multiplied by the percentage of their ratings that are mutual (one-sided).

Descriptive Statistics

Table 1 reports descriptive statistics for our sample.[16] Panel A contains descriptive statistics at the receiver-year level, Panel B at the relationship-formation level, and Panel C at the first stage rating level. On average, an employee received 4.87 first stage ratings for the period. MRRs are prevalent, with an employee receiving 2.27 mutual ratings on average.[17] Employees in our sample are split almost equally by gender, have an average tenure of 2.66 years, and the majority are less than 40 years old. 61% are natives of the country where the company is headquartered. The employees span six hierarchical levels, from 0 (lowest) to 5 (highest), with a median level of 3. The average supervisor rating in period 2 is 2.09 (the average post-calibration rating (untabulated)

[16] Appendix B details the inclusion criteria for each sample; Appendix C provides a reconciliation of nominations and actual ratings for period 2.
[17] Of our 7,442 employee-year observations, 1,229 have no mutual ratings and 536 have only mutual ratings.

is 2.05, with 2% (7%) of ratings adjusted upwards (downwards) by calibration committees).

Approximately 56% of rater nominations are mutual, i.e., the potential giver is nominated to rate the receiver and vice versa. Supervisors veto around 11% of nominations, and rater additions by supervisors lead to a mutual rating request 43% of the time. Less than 7% of rating requests are declined. The average rating received in the first stage is 3.56; the standard deviation is 0.78. The average mutual rating is 3.70, while the average one-sided rating is 3.45, providing initial evidence of a mutual rating premium ($p < 0.01$, two-tailed). Receivers and givers are demographically similar (have organizational proximity) in 32% (39%) of rating instances.

IV. EMPIRICAL TESTS AND RESULTS

Mutual Rating Relationship Formation

We examine the factors associated with the likelihood of a mutual rater nomination, specifically whether demographic similarity (H1a), organizational proximity (H1b), and strong cooperation incentives (H1c) increase the likelihood of a mutual versus one-sided nomination. We begin by analyzing whether these factors influence nominations in general. Then, to test our formal hypotheses, we examine if they explain the likelihood of a *mutual* nomination, given there is a nomination.

To examine nominations in general, we treat all identifiable employees in our sample for a given period (except the receiver's primary supervisor and the receiver themself) as potential "givers," i.e., all could potentially rate the receiver. We estimate the following model using an ordinary least squares (OLS) regression, to examine the likelihood of potential giver *j* being nominated to rate receiver *i* in period 2 (our nomination data pertains only to this period):

$$Nominee_{ij} = \beta_0 + \beta_1 Demographic\ Similarity_{ij} + \beta_2 Organizational\ Proximity_{ij} + \beta_3 Noncompetitive\ Promotion\ Prospects_{ij} + Receiver\ FEs + (Potential)\ Giver\ FEs + \varepsilon_{ij} \quad (1)$$

We report our results from estimating (1) in Table 2, column 1. We find that both demographic similarity and organizational proximity between a receiver and a potential giver increase the likelihood of the potential giver being nominated, as does strong cooperation incentives between the two (proxied by *Noncompetitive Promotion Prospects*). To examine if these variables explain *mutual* nominations, above and beyond their effect on nominations, we estimate the following model, again using an OLS regression:

$$\begin{aligned} \textit{Mutual Nomination}_{ij} = \quad & \beta_0 + \beta_1 \textit{ Demographic Similarity}_{ij} + \beta_2 \textit{ Organizational Proximity}_{ij} + \\ & \beta_3 \textit{ Noncompetitive Promotion Prospects}_{ij} + \textit{Receiver FEs} + \\ & (\textit{Potential) Giver FEs} + \varepsilon_{ij} \end{aligned} \tag{2}$$

Our sample comprises observations where *Nominee* equals 1, excluding those where the nominee is a direct subordinate of the receiver because, by definition, a mutual nomination cannot exist (a primary supervisor should not be nominated to rate their direct subordinate). We report our results from estimating (2) in Table 2, column 2. We find that a mutual rater nomination is more likely with demographic similarity or organizational proximity. The coefficient on *Noncompetitive Promotion Prospects* is positive (and larger than that on *Demographic Similarity*) but insignificant.[18] We thus find support for H1a and H1b, but not H1c. Demographic similarity (organizational proximity) increases the likelihood of a mutual nomination by approximately 2 (14) percentage points.

The formation of MRRs may be determined by both employee nominations and supervisor involvement in the relationship formation process. In our setting, supervisors could veto rater nominations in period 2 and add raters. We first examine if supervisor vetoing varies with the nature of the nomination under examination—mutual or one-sided. We then examine if the

[18] There are two possible explanations for this null result. First, our measure of strong cooperation incentives applies to less than four percent of observations, which may limit statistical power. Second, our definition requires that both the (potential) giver and receiver are considered "ready for promotion," which employees cannot necessarily predict ex ante.

likelihood of a supervisor rater addition creating a mutual (as opposed to a one-sided) rating request varies with the characteristics of the receiver-giver pair.

To explore whether supervisors are more or less likely to veto rater nominations that are expected to result in a MRR, we estimate the following model using an OLS regression:

$$\begin{aligned} \textit{Supervisor Vetoed} \\ \textit{Rater Nomination}_{ij} = \end{aligned} \quad \begin{aligned} &\beta_0 + \beta_1 \textit{Mutual Nomination}_{ij} + \beta_2 \textit{Demographic Similarity}_{ij} + \\ &\beta_3 \textit{Organizational Proximity}_{ij} + \beta_4 \textit{Noncompetitive Promotion} \\ &\textit{Prospects}_{ij} + \beta_5 \textit{R\_Nominations}_i + \textit{Controls} + \textit{Fixed Effects} + \varepsilon_{ij} \end{aligned} \quad (3)$$

At the receiver level, we control for gender (*R_Female*), nationality (*R_Native*), age category (the variables beginning with *R_Age*), company tenure (*R_Tenure*), and hierarchical rank (*R_Rank*). We control for these same variables at the giver level (we use the same naming conventions, replacing *R_* with *G_*), as well as whether the giver is a primary subordinate of the receiver (*G_Sub*). We include receiver and (proposed) giver department fixed effects to control for any department characteristics that may differentially affect the likelihood of a rater nomination being vetoed. We also control for *R_Nominations* under the assumption that the likelihood of a supervisor vetoing a nomination is increasing in the number of rater nominations a receiver has (the results confirm this). We present the results of this analysis in Table 3, columns 1 and 2. Column 1 (2) excludes (includes) receiver and giver fixed effects.

We find a negative and significant coefficient on *Mutual Nomination*, that is, supervisors are *less likely* to veto nominations that would result in a MRR. Nominations are more likely to be vetoed in the presence of demographic similarity between the receiver and potential giver (column 2 which is the more stringent specification), and less likely to be vetoed in the presence of organizational proximity (columns 1 and 2). It thus appears that supervisors have a preference to limit rating relationships that may result in more favorable evaluations due to demographic similarity but to enable relationships that may yield greater accuracy stemming from familiarity.

Strong cooperation incentives is unrelated to the likelihood of a nomination being vetoed.

Next, we examine supervisors' rater additions (i.e., the giver was not nominated, but the supervisor added the giver as a rater). Specifically, we examine whether the likelihood that an addition leads to a supervisor-created mutual (as opposed to a one-sided) rating request varies with demographic similarity, organizational proximity, or strong cooperation incentives between the receiver and giver, that is, whether it varies with the characteristics of the receiver-giver pair. To do so, we estimate the following model using an OLS regression:

$$\begin{aligned} \textit{Supervisor-Created Mutual Rating Request}_{ij} = {} & \beta_0 + \beta_1\, \textit{Demographic Similarity}_{ij} + \beta_2\, \textit{Organizational Proximity}_{ij} + \\ & \beta_3\, \textit{Noncompetitive Promotion Prospects}_{ij} + \beta_4\, \textit{R\_Approved Nominations to Rate}_i + \beta_5\, \textit{R\_Pct. Mutual Approved Nominations}_i + \\ & \textit{Controls} + \textit{Fixed Effects} + \varepsilon_{ij} \end{aligned} \quad (4)$$

The sample comprises supervisor rater additions, except those where the giver is a primary subordinate of the receiver (since by design, a mutual rating request cannot arise).[19] The receiver and giver controls are the same as when estimating (3) (except *G_Sub* is not applicable), as are the department fixed effects. We control for *R_Approved Nominations to Rate*, as the likelihood that a rater addition leads to a supervisor-created mutual rating request is likely increasing with the number of approved nominations that exist for the *receiver* to provide a rating, and we control for *R_Pct. Mutual Approved Nominations* as the percentage of existing approved mutual nominations may influence supervisor rater addition behavior. The results are reported in Table 3, column 3.

We find that the likelihood of the supervisor creating a mutual rating request is higher when there is organizational proximity between the receiver and giver, while neither demographic

[19] Typically, a supervisor-created mutual rating request arises where there is an existing approved nomination for the receiver to rate the giver, and a supervisor then adds the giver to rate the receiver. We limit our sample to receivers who have at least one approved nomination to rate an employee because supervisors' ability to create mutual rating requests critically depends on those nominations. Nonetheless, supervisors can also create a mutual rating request by adding both sides of the relationship (i.e., a supervisor adds the receiver to rate the giver, and a supervisor adds the giver to rate the receiver). While we do not exclude such additions from our sample, they are rare (7.06% of our sample in Table 3) and our results are robust to excluding such additions (untabulated).

similarity nor strong cooperation incentives affects the likelihood of a supervisor-created mutual rating request. Thus, supervisors appear to be willing to create a mutual rating relationship when there is likely greater familiarity between the receiver and giver. We find that the likelihood of the supervisor creating a mutual rating request is decreasing with the percentage of the receiver's approved nominations that are mutual.

Overall, we find an affirmative answer to RQ1: supervisors do impact MRR formation. Notably, supervisors are less likely to veto nominations that would result in a MRR all else equal; thus, supervisors contribute to the prevalence of MRRs through their vetoing behavior. However, supervisor rater additions more often create a one-sided rating request than a mutual rating request (see Table 1, Panel B, Subpanel 3 where the mean of *Supervisor-Created Mutual Rating Request* is 0.43) and their rater additions are less likely to create mutual (vis-à-vis one-sided) rating requests when receivers are expected to have a greater proportion of MRRs. Thus, supervisors appear to be somewhat cognizant of the prevalence of MRRs during the rater addition process.

In our research setting, employees who are asked to provide a rating can decline the request. To test our prediction that a designated giver will be less likely to decline a request that, if accepted, should result in a MRR, we estimate the following model using an OLS regression:

$$\begin{aligned} \textit{Rating Request Declined}_{ij} = {} & \beta_0 + \beta_1\, \textit{Mutual Approved Rating Request}_{ij} + \\ & \beta_2\, \textit{Demographic Similarity}_{ij} + \beta_3\, \textit{Organizational Proximity}_{ij} + \\ & \beta_4\, \textit{Noncompetitive Promotion Prospects}_{ij} + \beta_5\, \textit{G\_Requests to Rate}_{j} \\ & + \beta_6\, \textit{R\_Requests to be Rated}_{i} + \textit{Controls} + \textit{Fixed Effects} + \varepsilon_{ij} \end{aligned} \quad (5)$$

The variable of interest is *Mutual Approved Rating Request*, which captures the receiver and giver each being asked to rate the other (which may stem from approved nominations and/or supervisor rater additions), and we expect a negative coefficient. In addition to our standard receiver and giver controls, we control for the number of rating requests the designated giver received (*G_Requests to Rate*) as we expect the likelihood of a decline to be increasing in the number of requests. We

also control for the number of rating requests that were made for the receiver (*R_Requests to be Rated*), since to the extent that a giver is aware of how many requests were made, they may be more willing to decline a request the greater the number of total requests made.

We report the results of estimating (5) in Table 4, Panel A. We find a negative and significant coefficient on *Mutual Approved Rating Request*—employees are less likely to decline a rating request if accepting is expected to result in a MRR (supporting H2). Employees are less likely to decline a request in the presence of organizational proximity, suggesting their familiarity with the receiver influences their willingness to provide a rating, while neither demographic similarity nor strong cooperation incentives affects the likelihood of decline. Consistent with our expectations, we find positive and significant coefficients on *G_Requests to Rate* and *R_Requests to be Rated*.

Employees responding more favorably to individual rating requests that are expected to result in a MRR is consistent with MRRs capturing preferred rating relationships. We thus examine in a follow-up analysis whether an employee's general willingness to provide ratings varies with supervisor vetoing of their own mutual rater nominations. We estimate the following model using an OLS regression at the designated giver level in period 2:

$$\begin{aligned} \textit{Any Requests Declined}_i = {} & \beta_0 + \beta_1\, \textit{G\_Any Supervisor Vetoes-Mutual}_i + \beta_2\, \textit{G\_Any Supervisor Vetoes-One-sided}_i + \beta_3\, \textit{G\_Requests to Rate}_i + \beta_4\, \textit{G\_Pct. Mutual Nominations To Rate} + \textit{Controls}_i + \varepsilon_i \end{aligned} \quad (6)$$

If supervisor vetoing of mutual nominations in particular lowers employees' willingness to participate in the multi-rater system, we should find a positive coefficient on *G_Any Supervisor Vetoes-Mutual* and this should be higher than the coefficient on *G_Any Supervisor Vetoes-One-sided*. We control for the number of requests to rate at the giver level (*G_Requests to Rated*) and the percentage of the requests that are mutual (*G_Pct. Mutual Nominations To Rate*).

Table 4, Panel B, column 1, provides the results of this analysis. As expected, we find the coefficient on *G_Any Supervisor Vetoes-Mutual* is positive and significantly higher than that on *G_Any Supervisor Vetoes-One-sided* ($p < 0.01$). Table 4, Panel B, column 2 provides the results for a second specification investigating the percentage of requests declined (*Pct. Requests Declined*) as an alternative dependent variable and we find consistent results. In both columns we find that employees are more likely to decline rating requests when they receive a higher number of overall requests (*G_Requests To Rate*), but notably less likely to do so when a greater proportion of those requests involve *mutual* nominations (*G_Pct. Mutual Nominations To Rate*). This provides further evidence that MRRs reflect employees' preferred rating relationships. Overall, the results show that employees are more inclined to decline rating requests when their own mutual nominations are vetoed by the supervisor. This pattern may help explain why supervisors are more reluctant to veto mutual nominations (vis-à-vis one-sided nominations), as shown in Table 3, columns 1 and 2.

Finally, we examine *Demographic Similarity, Organizational Proximity* and *Strong Cooperation Incentives* as determinants of MRRs, that is, the final relationships after supervisor vetoes and additions, and employee declines. This analysis is conducted for both period 2—the time frame used in the preceding tables—and the pooled sample covering both periods (with period fixed effects included in the latter). The results, presented in Table 5 show that *Demographic Similarity*, *Organizational Proximity* and in three out of four specifications, *Strong Cooperation Incentives*, are positively and significantly associated with MRR formation. *Organizational Proximity* exerts the strongest influence of these three determinants.

Mutual Rating Relationship Outcomes

We now examine whether mutual ratings differ systematically from one-sided ratings

(RQ2), that is, if they are higher or lower on average. We estimate the following model with an OLS regression using the pooled sample (i.e., period 1 and period 2 data):

$$\begin{aligned} \textit{Rating Received}_{ijt} = \; & \beta_0 + \beta_1 \textit{Mutual Rating Relationship}_{ijt} + \beta_2 \textit{Demographic Similarity}_{ijt} \\ & + \beta_3 \textit{Organizational Proximity}_{ijt} + \beta_4 \textit{Noncompetitive Promotion Prospects}_{ijt} + \beta_5 \textit{R\_Mean Rating Received}_{it} + \beta_6 \textit{G\_Mean Rating Given}_{jt} + \textit{Controls} + \textit{Fixed Effects} + \varepsilon_{ijt} \end{aligned} \quad (7)$$

The primary variable of interest is *Mutual Rating Relationship*. In addition to our standard set of receiver and giver controls, we include *R_Mean Rating Received* to control for the receiver's performance and *G_Mean Rating Given* to control for potential variation in rating tendencies across givers (e.g., leniency in ratings) that is not captured by demographic characteristics, rank, or department, all of which we control for separately (see Appendix A for the specific variable definitions).[20] We include period fixed effects to control for any period differences, and receiver and giver department fixed effects to control for any time-invariant department characteristics that may affect ratings. We first estimate (7) without *Demographic Similarity*, *Organizational Proximity*, and *Noncompetitive Promotion Prospects* as these are determinants of *Mutual Rating Relationship* as shown in Table 5. For each estimation, we include receiver fixed effects and giver fixed effects in one specification, which control for time-invariant characteristics that could affect the ratings an employee receives or the ratings an employee gives. Our results are reported in Table 6, Panel A, columns 1 through 4.

In all columns, we find a positive and significant coefficient on *Mutual Rating Relationship*—thus, mutual ratings are higher on average than one-sided ratings. The coefficients indicate that mutual ratings are between 0.187 to 0.212 scale points higher on average, equivalent

[20] We include *G_Sub* as a control because, by design, a MRR cannot exist in this instance, and because we expect (and find; not reported) higher ratings from subordinates due to power distance (Atwater, Brett, and Charles 2007).

to a quarter of a standard deviation. We thus find evidence of a mutual rating "premium."[21] *Demographic Similarity* is associated with higher ratings (consistent with greater liking), as is *Noncompetitive Promotion Prospects*, while *Organizational Proximity* is associated with lower ratings (perhaps due to competition between employees). Controlling for these determinants does not dampen the magnitude of the mutual rating premium. As expected, we find a positive and significant coefficient on *R_Mean Rating Received* (the higher the average rating received by the receiver, the higher the current rating).

In columns 5 and 6, we interact *Mutual Rating Relationship* with each of *Demographic Similarity*, *Organizational Proximity*, and *Noncompetitive Promotion Prospects*. We find a much more pronounced mutual rating premium in the presence of strong cooperation incentives (i.e., the interaction with *Noncompetitive Promotion Prospects*), consistent with strategic rating behavior. We find some evidence of a slightly larger premium when there is organizational proximity between the receiver and giver, while there is no evidence that the magnitude of the premium varies with demographic similarity.

Does Strategic Rating Behavior Influence the Mutual Rating Premium?

Table 6, Panel A provides initial evidence of strategic rating behavior in MRRs by identifying a more pronounced mutual rating premium when there are strong cooperation incentives between the receiver and giver. We design two additional empirical tests to investigate whether there is further evidence of strategic rating behavior in MRRs. Our first test relies on the rater nomination data, which identifies supervisor vetoes. Thus, in addition to ratings from actual

[21] In untabulated analyses, we replace *Mutual Rating Relationship* with a set of indicator variables to capture the specific rating the receiver gives to the giver. We find a monotonic relationship between the rating given in a MRR and the rating received. Thus, rather than a constant premium, the mutual rating premium on *Rating Received* increases with the reverse rating (i.e., the rating given). We do not find a significant mutual rating premium above one-sided ratings for the receiver when the rating given is 1 or 2.

MRRs, we can identify ratings from proposed MRRs where the reverse nomination was vetoed (*Proposed MRR Vetoed*), that is, ratings from givers where the receiver would have also rated the giver, had the relevant nomination *not* been vetoed. When a proposed MRR is vetoed, any rating premium from selection effects (such as interpersonal attraction or a favorable shared working experience) should remain even in a one-sided rating relationship. However, any incentives to engage in strategic rating behavior should be muted in the remaining one-sided rating relationship.[22] Consequently, if the coefficient on *Mutual Rating Relationship* is significantly larger than the coefficient on *Proposed MRR Vetoed*, there is evidence of strategic rating behavior contributing to the mutual rating premium. Note that this test does not require that supervisors veto relationships randomly. If anything, we expect supervisors will be more likely to veto a nomination where they expect a resulting rating would provide an overly favorable assessment of performance. This would work against finding a difference in the coefficients on *Proposed MRR Vetoed* and *Mutual Rating Relationship*.[23]

We present the results of this analysis in columns 1 (no receiver or giver fixed effects) and 2 (with receiver and giver fixed effects) of Table 6, Panel B. We find positive and significant coefficients on *Mutual Rating Relationship* and *Proposed MRR Vetoed*. More importantly, the coefficient on *Mutual Rating Relationship* is statistically significantly higher than that on *Proposed MRR Vetoed*, suggesting that at least part of the premium is driven by strategic rating behavior.

Our second test is based on economic theory outlined in the development of RQ3 that employees may be less likely to engage in strategic rating behavior if the MRR is unlikely to continue. To the extent that employees can anticipate which MRRs are more likely to persist

[22] There may still be some incentive to engage in strategic rating behavior if the giver expects the receiver may rate them in a future period, but any such incentive should be smaller relative to a MRR in the current period.
[23] If, for some reason, some raters are unaware that the nomination for the receiver to rate them has been or will be vetoed at the time they give their rating, this will also work against finding a difference.

beyond the current period, we expect any strategic rating behavior to reflect their belief in MRR persistence. First, we use the period 1 data and run a regression that splits *Mutual Rating Relationship* into two indicators, *MRR Both Periods* and *MRR Period 1 Only*. If *MRR Both Periods* reflects a stronger (correct) employee belief in MRR persistence, we expect to find a positive coefficient on *MRR Both Periods* that is significantly larger than the (positive) coefficient on *MRR Period 1 Only*. That is, we use the presence or absence of a continued MRR in period 2 to proxy for an employee's ex-ante beliefs regarding MRR persistence in period 1, which we expect to influence the premium in period 1. We assume that the types of interactions or relationships that employees have may allow them to distinguish between rating relationships that are a one-time MRR occurrence and those that are more likely to give rise to persistent MRRs. Second, we use the period 2 data and run a regression that splits *Mutual Rating Relationship* into *MRR Both Periods* and *MRR Period 2 Only* (we also include *MRR Period 1 Only*). Again, if strategic rating behavior is at play, we expect a positive coefficient on *MRR Period 2 Only* because of current strategic rating incentives. However, we expect the coefficient to be smaller in magnitude than the coefficient on *MRR Both Periods* (the premium here should be most pronounced because of the continued rating relationship that incentivizes quid-pro-quos). If *MRR Period 1 Only* reflects a prior MRR that may or may not be repeated in the future, we may find a positive coefficient, but one that is likely significantly lower than the coefficients on *MRR Period 2 Only* and *MRR Both Periods*. In sum, if strategic rating behavior contributes to the mutual rating premium, we expect *MRR Both Periods* > *MRR Period 2 Only* > *MRR Period 1 Only*. We consider these analyses (i.e., the one using period 1 rating data and the one using period 2 rating data) as complementary as they use different measures to capture beliefs in MRR persistence.

We present the results of these analyses in Table 6, Panel C. In columns 1 and 2, we find

positive and significant coefficients on *MRR Both Periods* and *MRR Period 1 Only*, and the coefficient on the former is statistically significantly higher than the coefficient on the latter. Thus, where we expect employees to have a stronger belief that the MRR will hold in future periods, the mutual rating premium in the current period is more pronounced. In columns 3 and 4, we find a positive and significant coefficient on both *MRR Both Periods* and *MRR Period 2 Only*, with the coefficient on the former statistically significantly larger than the coefficient on the latter. We also find the coefficient on *MRR Period 1 Only* is positive and significant, and, as expected, statistically significantly smaller than *MRR Period 2 Only* and *MRR Both Periods.*

Collectively, our tests across all three panels of Table 6 show that there are conditions under which strategic behavior arises and partly contributes to the documented mutual rating premium (rather than the premium being solely driven by selection in relationship formation).

Supervisor Rating Behavior and Mutual Rating Relationships

We now turn to RQ4, on whether supervisors place differential weight on mutual and one-sided ratings when evaluating employees. Supervisor behavior during the relationship formation process may influence their later rating behavior (e.g., due to vetoing or rater additions) and we can only examine the relationship formation process for period 2. Therefore, we restrict our sample to period 2 data and estimate the following OLS regression with one observation per receiver *i*:

$$\begin{aligned} Supervisor\ Rating_i = \quad & \beta_0 + \beta_1\, R\_Weighted\ Mean\ Mutual\ Rating_i + \\ & \beta_2\, R\_Weighted\ Mean\ One\text{-}sided\ Rating_i + \\ & \beta_3\, R\_Pct.\ Nominations\ Vetoed_i + \beta_4\, R\_All\ Raters\ Supervisor \\ & Additions_i + Controls + Fixed\ Effects + \varepsilon_i \end{aligned} \tag{8}$$

We include the standard set of receiver controls, a control for the percentage of nominations that were vetoed (*R_Pct. Nominations Vetoed*), a control capturing whether all raters for a receiver were added by the supervisor (*R_All Raters Supervisor Additions*), receiver department fixed

effects, and in some specifications, supervisor fixed effects.[24] We report our results in Table 7, columns 1 (no supervisor fixed effects) and 2 (with supervisor fixed effects).

If supervisors weight mutual and one-sided ratings equally, then the coefficient β_1 will not be statistically different from the coefficient β_2. We find β_1 is statistically significantly larger than β_2 in column 1 ($p = 0.055$), but not in column 2 at conventional significance levels ($p = 0.113$). Thus, we find some evidence that the implicit weight supervisors place on mutual ratings is higher than that placed on one-sided ratings, suggesting that supervisors may perceive mutual ratings to be more informative than one-sided ratings when determining their employee evaluations.

We next examine whether any differential weighting of ratings depends on the number of ratings available for a given employee. Table 7, column 3 (4) comprises supervisor ratings for employees with less than five ratings (five or more ratings) in our sample (this represents a median split). We find strong evidence that supervisors place more weight on mutual ratings than on one-sided ratings when they have relatively fewer ratings available, but no evidence of differential weighting when they have relatively more ratings.[25] This finding is consistent with two possible explanations which are not mutually exclusive. First, when information is limited, one-sided ratings may be perceived as noisier. In such cases, supervisors might anchor more strongly on mutual ratings, which reflect reciprocal relationships and may signal greater interpersonal familiarity and, thus, perceived reliability. Second, when supervisors are confronted with a more demanding evaluation task, i.e., more ratings to consider, they may simplify their evaluation

[24] We include *R_All Raters Supervisor Additions* as it is rare for an employee to have all their raters in our sample result purely from supervisor additions (3.39% of the sample in Table 7). In most cases, none of the raters for a given employee results from supervisor additions (77.53% of the sample).

[25] In an untabulated analysis, we explore the possibility that when supervisors use their discretion to impact the employee's rating relationships (i.e., vetoing nominations or adding raters) they might also be more likely to differentiate between mutual and one-sided ratings. We split our sample into employees where the supervisor used such discretion and employees where the supervisor did not. Using the specification as in Table 7, column 2, we find no significant differences in either subsample.

process by assigning uniform weights to ratings, foregoing the nuanced differentiation between mutual and one-sided evaluations.[26]

Our major finding is that supervisors appear to consider mutual ratings as more informative when they have relatively fewer ratings available for their performance evaluation. Such greater perceived informativeness of mutual ratings is consistent with supervisors' lower willingness to veto mutual nominations vis-à-vis one-sided ones. In an effort to understand why (at least in some situations) supervisors may perceive mutual ratings to be more informative, we investigate in a separate analysis (untabulated) the length of the textual comments accompanying provided ratings. We find that the comments accompanying mutual ratings are, on average, longer than those for one-sided ratings ($p < 0.01$ when performing a t-test of the difference in means of the number of words). To the extent that longer comments provide supervisors with more information, these results align with supervisors placing more weight on mutual ratings (in some cases).[27]

Recall that the calibration committee has the final say on ratings. In untabulated analyses, we repeat column 1 of Table 7 replacing *Supervisor Rating* with *Committee Rating* (and including *Supervisor Rating*). We find a positive and significant coefficient on both *R_Weighted Mean Mutual Rating* and *R_Weighted Mean One-sided Rating* but no significant difference. These results suggest that calibration committees consider both equally informative beyond the supervisor rating and, in line with prior work, supervisors do not have full decision authority over employee performance evaluation (Demere et al. 2019).

[26] A concurrent paper (Bol et al. 2023) examines how supervisor weighting of multi-rater ratings varies with what the authors refer to as the supervisor's "cognitive load", but it does not consider mutual and one-sided ratings.

[27] Scholars have highlighted supervisors' aversion to providing negative feedback to employees (Harris 1994; Beer 1997; Prendergast 1999). Thus, another reason why supervisors may be less likely to veto mutual nominations is that the comments for such ratings might serve as an information source and inspiration for positive feedback that supervisors can share in one-on-one meetings with their employees. Indeed, we find that the comments accompanying mutual ratings have a more positive sentiment than those that accompany one-sided ones ($p < 0.01$ for mean difference). To ensure comparability of sentiment scores, we restrict the analysis to comments in English (62% of comments; the remainder are in the native language of the headquarters' country).

V. CONCLUSION

This paper examines MRRs, which are likely to be prevalent in any multi-rater system. Using proprietary data from an online retailer, we find that mutual rater nominations are more likely with demographic similarity or organizational proximity between employees. Supervisors are less likely to veto rater nominations that would lead to a MRR, but they appear more cautious about actively creating additional MRRs through their own rater additions. We find that employees are more likely to respond favorably to rating requests when these are expected to result in a MRR, and they are generally more willing to provide ratings when supervisors do not veto their own mutual nominations. Thus, preserving employee participation and satisfaction with the multi-rater system may be one potential contributing factor to supervisors' vetoing behavior.

There is a mutual rating premium, and this premium is more pronounced (consistent with strategic rating behavior playing a role) when employees in a MRR are more likely to believe in relationship persistence and when employees have strong incentives to cooperate for higher ratings. Supervisors place more weight on mutual ratings vis-à-vis one-sided ratings when they have relatively fewer ratings available for a given employee, indicating that supervisors may perceive mutual ratings to be more informative when there is limited information, despite any selection effects or strategic rating behavior.

Our study is subject to limitations. First, while prior work motivates our examination of demographic similarity and organizational proximity, we lack direct measures of employees' social ties and formal work relationships. Future research could examine relationships more directly by using internal records of employees' communications such as messaging or email exchanges (Mahlendorf, Martin, and Smith 2023). Second, though our results provide evidence consistent with strategic rating behavior, we lack a benchmark of how an employee's performance

is truly perceived. Thus, we cannot completely disentangle mutual ratings that are purposely inflated from those that are not. However, a reason why firms establish multi-rater systems is the lack of objective measures providing such a benchmark. Third, our analyses are limited to a single firm and thus may not generalize to all settings. Nonetheless, the system we study is representative of contemporary multi-rater systems according to survey evidence and the practitioner literature.

Despite its limitations, our study increases our understanding of how multi-rater systems operate in practice. Our findings provide insights relevant to designers and users when considering the benefits and costs of designing, implementing, and using such systems. The findings suggest that mutual rating relationships matter, as we observe systematic differences between mutual and one-sided ratings, and we also find evidence that supervisors respond to mutual rating relationships and ratings differently (at least in some cases) vis-à-vis one-sided ones. However, since we examine on-average effects, there may be a non-trivial proportion of supervisors who are not necessarily aware of or responsive to different relationship types. Firms may benefit from making the type of relationship transparent to supervisors throughout the relationship formation and rating processes and also to calibration committees, as allowing users of multi-rater systems to readily have access to this information could help foster more informed decision-making processes.

REFERENCES

Aadland, E., G. Cattani, and S. Ferriani. 2019. Friends, gifts, and cliques: Social proximity and recognition in peer-based tournament rituals. *Academy of Management Journal*, *62(3)*, 883-917.

Abeler, J., D. Nosenzo, and C. Raymond. 2019. Preferences for truth-telling. *Econometrica*, *87(4)*, 1115-1153.

Antonioni, D., and H. Park. 2001. The relationship between rater affect and three sources of 360-degree feedback ratings. *Journal of Management*, *27(4)*, 479-495.

Atwater, L. E., J. F. Brett, and A. C. Charles. 2007. Multisource feedback: Lessons learned and implications for practice. *Human Resource Management*, *46(2)*, 285-307.

Badura, K. L., E. Grijalva, D. A. Newman, T. T. Yan, and G. Jeon. 2018. Gender and leadership emergence: A meta-analysis and explanatory model. *Personnel Psychology*, *71*, 335-367.

Beer, M. 1997. Conducting a performance appraisal interview. Harvard Business School Background Note 497-058.

Berger, J., C. Harbring, and D. Sliwka. 2013. Performance appraisals and the impact of forced distribution—An experimental investigation. *Management Science*, *59(1)*, 54-68.

Bol, J. C. 2008. Subjectivity in compensation contracting. *Journal of Accounting Literature*, *27*, 1-24.

Bol, J. C. 2011. The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review*, *86(5)*, 1549-1575.

Bol, J. C., M. Margolin, and D. Schaupp. 2023. Multi-rater performance evaluation and calibration: managing multiple opinions. Working paper.

Brutus, S., E. Aucoin, and S. Petosa. 2005. Who will evaluate me? Rater selection in multi-source assessment contexts. *International Journal of Selection and Assessment*, *13(2)*, 129-138.

Campbell, D. 2008. Nonfinancial performance measures and promotion-based incentives. *Journal of Accounting Research*, *46(2)*, 297-332.

Cappelli, P., and M. J. Conyon. 2018. What do performance appraisals do? *ILR Review*, *71(1)*, 88-116.

Cardinaels, E., and C. Feichter. 2021. Forced rating systems from employee and supervisor perspectives. *Journal of Accounting Research, 59(5)*, 1573-1607.

Conway, J. M., K. Lombardo, and K. C. Sanders. 2001. A meta-analysis of incremental validity and nomological networks for subordinate and peer ratings. *Human Performance*, *14(4)*, 267-303.

Dal Bó, P., and G. R. Fréchette. 2018. On the determinants of cooperation in infinitely repeated games: A survey. *Journal of Economic Literature*, *56(1)*, 60-114.

Deller, C. 2023. Beyond performance: Does assessed potential matter to employees' voluntary departure decisions? *Journal of Accounting Research*, *61(4)*, 981-1024.

Deller, C., S. Gallani, and T. Sandino. 2024. Promoting core values with a values-based performance assessment system: Evidence from a retail chain. Working paper.

Demere, B. W., K. L. Sedatole, and A. Woods. 2019. The role of calibration committees in subjective performance evaluation systems. *Management Science*, *65(4)*, 1562-1585.

Di Fiore, A., and M. Souza. 2021. Are peer reviews the future of performance evaluations? *Harvard Business Review*.

Du, F., Erkens, D., Young, M., and G. Tang. 2018. How adopting new performance measures affects subjective performance evaluations: evidence from EVA adoption by Chinese state-owned enterprises. *The Accounting Review, 93(1)*, 161-185.

Du, F., Tang, G., and M. Young. 2012. Influence activities and favoritism in subjective performance evaluation: Evidence from Chinese state-owned enterprises. *The Accounting Review*, 87 (5), 1555-1588.

Evans, J., L. Hannan, R. Krishnan, and D. Moser. 2001. Honesty in managerial reporting. *The Accounting Review*, *76(4)*, 537-559.

Frederiksen, A., L. B. Kahn, and F. Lange. 2020. Supervisors and performance management systems. *Journal of Political Economy*, *128(6)*, 2123-2187.

Grabner, I., J. Künneke, and F. Moers. 2020. How calibration committees can mitigate performance evaluation bias: An analysis of implicit incentives. *The Accounting Review*, *95(6)*, 213-233.

Grabner, I., and F. Moers. 2013. Managers' choices of performance measures in promotion decisions: An analysis of alternative job assignments. *Journal of Accounting Research*, *51(5)*, 1187-1220.

Harbring, Ch. and B. Irlenbusch. 2008. How many winners are good to have?: On tournaments with sabotage. *Journal of Economic Behavior & Organization, 65(3–4)*, 682-702.

Harris, M. M. 1994. Rater motivation in the performance appraisal context: A theoretical framework. *Journal of Management*, *20(4)*, 737-756.

Hecht, G., V. S. Maas, and M. van Rinsum. 2023. The effects of transparency and group incentives on managers' strategic promotion behavior. *The Accounting Review*, *98(7)*, 239-260.

Huang, Y., M. Shum, X. Wu, and J. Z. Xiao. 2019. Discovery of bias and strategic behavior in crowdsourced performance assessment. Working paper.

Ibarra, H. 1992. Homophily and differential returns: Sex differences in network structure and access in an advertising firm. *Administrative Science Quarterly, 37(3)*, 422-447.

Ittner, C. D., D. F. Larcker, and M. W. Meyer. 2003. Subjectivity and the weighting of performance measures: Evidence from a balanced scorecard. *The Accounting Review*, *78(3)*, 725-758.

Kiessling, L., J. Radbruch, and S. Schaube. 2019. Determinants of peer selection. CRC TR 224 Discussion Paper No. 092, University of Bonn and University of Mannheim.

Kiessling, L., J. Radbruch, and S. Schaube. 2022. Self-selection of peers and performance. *Management Science*, *68(11)*, 7793-8514.

Kingstrom, P. O., and L. E. Mainstone. 1985. An investigation of the rater-ratee acquaintance and rater bias. *Academy of Management Journal, 28(3)*, 641-653.

Klapper, H., H. Piezunka, and L. Dahlander. 2024. Peer evaluations: Evaluating and being evaluated. *Organization Science,* 35(4), 1363-1387.

Künneke, J. 2024. Supervisor impact on employee careers: The role of rating differentiation. *Accounting Review*, 99(6), 271–295.

Lawrence, B. S., and N. P. Shah. 2020. Homophily: Measures and meaning. *Academy of Management Annals, 14(2)*, 513-597.

Lipe, M. G., and S. E. Salterio. 2000. The balanced scorecard: Judgmental effects of common and unique performance measures. *The Accounting Review, 75(3)*, 283-298.

London, M., and J. W. Smither. 1995. Can multi-source feedback change perceptions of goal accomplishment, self-evaluations, and performance-related outcomes? Theory-based applications and directions for research. *Personnel Psychology*, *48(4)*, 803-839.

Luft, J. 2016, Cooperation and competition among employees: Experimental evidence on the role of management control systems. *Management Accounting Research, 31*, 75-85.

Mahlendorf, M., M. Martin and D. A. Smith. 2023. Innovative data – Use-cases in management accounting research and practice. *European Accounting Review, 32(3),* 547-576.

Marsen, P. V. 1988. Homogeneity in confiding relations. *Social Networks, 10*, 57-76.

Milgrom, P., and J. Roberts. 1992. Economics, organization and management. Englewood Cliffs, NJ: Prentice Hall International.

Norton, S. M. 1992. Peer assessments of performance and ability: An exploratory analysis of statistical artifacts and contextual moderators. *Journal of Business and Psychology*, *6(3)*, 387-399.

Pfeffer, J. 1985. Organizational demography: Implications for management. *California Management Review*, *28(1)*, 67-81.

Prendergast, C. 1999. The provision of incentives in firms. *Journal of Economic Literature*, *37(1)*, 7-63.

Prendergast, C. 2002. Uncertainty and incentives. *Journal of Labor Economics*, *20(S2)*, S115-S137.

Prendergast, C., and R. Topel. 1996. Favoritism in organizations. *The Journal of Political Economy*, *104(5)*, 958-978.

Raub, W., V. Buskens, and V. Frey. 2019. Strategic tie formation for long-term exchange relations. *Rationality and Society, 31(4)*, 490-510.

Riedl, C., T. Grad, and C. Lettl. 2024. Competition and collaboration in crowdsourcing communities: What happens when peers evaluate each other?. *Organization Science, 35(6),* 1957-1976.

Saal, F. E., and F. J. Landy. 1977. The mixed standard rating scale: An evaluation. *Organizational Behavior and Human Performance, 18(1)*, 19-35.

Shrum, W., N. H. Cheek, and S. M. Hunter. 1988. Friendship in school: Gender and racial homophily. *Sociology of Education, 61(4)*, 227-239.

Simons, R. L., and C. Deller. 2022a. *Henkel: Building a Winning Culture (C)*. Harvard Business School Case No. 9-123-034. Harvard Business School Publishing: Boston.

Simons, R. L., and C. Deller. 2022b. *Henkel: Building a Winning Culture (D)*. Harvard Business School Case No. 9-122-055. Harvard Business School Publishing: Boston.

Simons, R., and N. Kindred. 2012. *Henkel: Building a Winning Culture*. Harvard Business School Case No. 9-112-060. Harvard Business School Publishing: Boston.

Sol, J. 2016. Peer evaluation: Incentives and coworker relations. *Journal of Economics & Management Strategy*, *25(1)*, 56-76.

Speckbacher, G., and M. Wiernsperger. 2022. Peer assessments in diverse teams: The role of team performance information. Working paper.

Stein, J. C. 1989. Efficient capital markets, inefficient firms: A model of myopic corporate behavior. *The Quarterly Journal of Economics*, *104*, 655-669.

Towry, K. L. 2003. Control in a teamwork environment—The impact of social ties on the effectiveness of mutual monitoring contracts. *The Accounting Review 78(4)*, 1069–1095.

Tsui, A. S., and B. Barry. 1986. Interpersonal affect and rating errors. *The Academy of Management Journal, 29(3)*, 586-599.

Tsui, A. S., and C. A. O'Reilly. 1989. Beyond simple demographic effects: The importance of relational demography in superior-subordinate dyads. *The Academy of Management Journal*, *32(2)*, 402-423.

Wang, X. M., K. F. E. Wong, and J. Y. Y. Kwong. 2010. The roles of rater goals and ratee performance levels in the distortion of performance ratings. *Journal of Applied Psychology*, *95(3)*, 546-561.

Wong, K. F. E., and J. Y. Y. Kwong. 2007. Effects of rater goals on rating patterns: Evidence from an experimental field study. *Journal of Applied Psychology*, *92(2)*, 577-585.

WorldatWork in collaboration with J. Bol, I. Grabner, R. Grasser, S. Loftus, T. Sandino, and K. Sedatole. 2019. Contemporary Performance Evaluation Practices.

Zhang, Y. 2008. The effects of perceived fairness and communication on honesty and collusion in a multi-agent setting. *The Accounting Review*, *83(4)*, 1125-1146.

Zenger, T. R., and B. S. Lawrence. 1989. Organizational demography: The differential effects of age and tenure distributions on technical communication. *The Academy of Management Journal*, *32(2)*, 353-376.

Table 1
Summary statistics
This table presents descriptive statistics at the receiver-year level (Panel A), relationship-formation level (Panel B), and first stage rating level (Panel C). All variables are defined in Appendix A. In Panel A, the descriptive statistics for *R_Weighted Mean Mutual Rating*, *R_Weighted Mean One-sided Rating*, *R_Pct. Nominations Vetoed*, *R_All Raters Supervisor Additions*, and *Supervisor Rating* are for period 2; the descriptive statistics for all other variables are for the pooled sample (i.e., periods 1 and 2). In Panel B, the descriptive statistics for the variables reported in subpanels 1-5 are for period 2; those reported in subpanel 6 are for the pooled sample. In Panel C, the descriptive statistics are for the pooled sample unless noted otherwise. † indicates the sample excludes observations where *G_Sub*=1 (i.e., the giver is a primary subordinate of the receiver). + Supervisor data is more sparsely populated in our sample than department and rank data. To maximize our sample, we only require supervisor data to be non-missing if it would impact the determination of *Organizational Proximity*. Thus, we have less observations for *Same Supervisor* than *Same Department* or *Same Rank*. Any differences between the number of observations in this table and those used in our multivariate analyses are due to singletons.

	N	Mean	P50	SD	Min	Max
Panel A. Receiver-year level						
Number of Ratings	7,442	4.872	5.000	2.093	1.000	20.00
Number of Mutual Ratings	7,442	2.267	2.000	1.770	0.000	12.00
R_Female	7,442	0.478	0.000	0.500	0.000	1.000
R_Native	7,442	0.610	1.000	0.488	0.000	1.000
R_Age <=30	7,442	0.303	0.000	0.460	0.000	1.000
R_Age 30-40	7,442	0.567	1.000	0.495	0.000	1.000
R_Age 40-50	7,442	0.105	0.000	0.306	0.000	1.000
R_Age 50+	7,442	0.025	0.000	0.155	0.000	1.000
R_Tenure	7,442	2.658	2.084	1.914	0.000	12.16
R_Rank	7,442	2.655	3.000	1.212	0.000	5.000
R_Weighted Mean Mutual Rating	3,334	1.936	2.000	1.067	0.000	4.800
R_Weighted Mean One-sided Rating	3,334	1.603	1.571	0.948	0.000	5.000
R_Pct. Nominations Vetoed	3,334	0.090	0.000	0.164	0.000	1.000
R_All Raters Supervisor Additions	3,334	0.034	0.000	0.181	0.000	1.000
Supervisor Rating	3,334	2.095	2.000	0.424	1.000	3.000
Panel B. Relationship-formation level						
Subpanel 1. (Potential) nomination level						
Nominee (potential noms.)	17,405,889	0.002	0.000	0.040	0.000	1.000
Demographic Similarity (potential noms.)	17,405,889	0.174	0.000	0.380	0.000	1.000
Organizational Proximity (potential noms.)	17,405,889	0.025	0.000	0.157	0.000	1.000
Noncompetitive Promotion Prospects (potential noms.)	17,405,889	0.035	0.000	0.183	0.000	1.000
Mutual Nomination† (actual noms.)	26,640	0.562	1.000	0.496	0.000	1.000
Demographic Similarity† (actual noms.)	26,640	0.313	0.000	0.464	0.000	1.000
Organizational Proximity† (actual noms.)	26,640	0.373	0.000	0.483	0.000	1.000
Noncompetitive Promotion Prospects† (actual noms.)	26,640	0.039	0.000	0.193	0.000	1.000
Subpanel 2. Nomination level						
Supervisor Vetoed Rater Nomination	28,199	0.113	0.000	0.316	0.000	1.000
Mutual Nomination	28,199	0.531	1.000	0.499	0.000	1.000
Demographic Similarity	28,199	0.311	0.000	0.463	0.000	1.000
Organizational Proximity	28,199	0.360	0.000	0.480	0.000	1.000
Noncompetitive Promotion Prospects	28,199	0.039	0.000	0.193	0.000	1.000
R_Nominations	28,199	8.762	8.000	3.642	1.000	32.00
Subpanel 3. Rater-addition level						
Supervisor-Created Mutual Rating Request†	1,629	0.430	0.000	0.495	0.000	1.000
R_Pct. Mutual Approved Nominations†	1,629	0.422	0.429	0.335	0.000	1.000
Demographic Similarity†	1,629	0.278	0.000	0.448	0.000	1.000
Organizational Proximity†	1,629	0.428	0.000	0.495	0.000	1.000
Noncompetitive Promotion Prospects†	1,629	0.029	0.000	0.167	0.000	1.000
R_Approved Nominations To Rate†	1,629	6.732	6.000	4.290	1.000	28.00

	N	Mean	P50	SD	Min	Max
Subpanel 4. Rating-request level						
Rating Request Declined	27,089	0.066	0.000	0.249	0.000	1.000
Mutual Approved Rating Request	27,089	0.498	0.000	0.500	0.000	1.000
Demographic Similarity	27,089	0.308	0.000	0.461	0.000	1.000
Organizational Proximity	27,089	0.370	0.000	0.483	0.000	1.000
Noncompetitive Promotion Prospects	27,089	0.038	0.000	0.192	0.000	1.000
G_Requests To Rate	27,089	9.371	8.000	4.696	1.000	32.00
R_Requests To Be Rated	27,089	7.969	7.000	3.150	1.000	24.00
Subpanel 5. Designated giver level						
Any Requests Declined	3,952	0.246	0.000	0.431	0.000	1.000
G_Any Supervisor Vetoes-Mutual	3,952	0.188	0.000	0.391	0.000	1.000
G_Any Supervisor Vetoes-One-sided	3,952	0.227	0.000	0.419	0.000	1.000
Pct. Requests Declined	3,952	0.045	0.000	0.098	0.000	1.000
G_Pct. Supervisor Vetoes-Mutual	3,952	0.068	0.000	0.166	0.000	1.000
G_Pct. Supervisor Vetoes-One-sided	3,952	0.104	0.000	0.222	0.000	1.000
G_Requests To Rate	3,952	7.071	6.000	4.042	1.000	32.00
G_Pct. Mutual Nominations To Rate	3,952	0.608	0.600	0.241	0.000	1.000
Subpanel 6. Final rating relationship level						
Mutual Rating Relationship†	33,341	0.490	0.000	0.500	0.000	1.000
Demographic Similarity†	33,341	0.319	0.000	0.466	0.000	1.000
Organizational Proximity†	33,341	0.405	0.000	0.491	0.000	1.000
Noncompetitive Promotion Prospects†	33,341	0.038	0.000	0.190	0.000	1.000
Panel C. First stage rating level						
Rating Received	35,001	3.562	4.000	0.775	1.000	5.000
Rating Received (Mutual ratings)	16,139	3.697	4.000	0.754	1.000	5.000
Rating Received (One-sided ratings)	18,862	3.447	3.000	0.774	1.000	5.000
Mutual Rating Relationship	35,001	0.461	0.000	0.498	0.000	1.000
Demographic Similarity	35,001	0.317	0.000	0.465	0.000	1.000
Same Nationality	35,001	0.686	1.000	0.464	0.000	1.000
Same Gender	35,001	0.647	1.000	0.478	0.000	1.000
Same Age	35,001	0.484	0.000	0.500	0.000	1.000
Same Tenure	35,001	0.222	0.000	0.416	0.000	1.000
Organizational Proximity	35,001	0.390	0.000	0.488	0.000	1.000
Same Supervisor+	33,914	0.343	0.000	0.475	0.000	1.000
Same Department	35,001	0.485	0.000	0.500	0.000	1.000
Same Rank	35,001	0.411	0.000	0.492	0.000	1.000
Noncompetitive Promotion Prospects	35,001	0.038	0.000	0.191	0.000	1.000
R_Mean Rating Received	35,001	3.570	3.600	0.493	1.000	5.000
G_Mean Rating Given	35,001	3.550	3.545	0.455	1.000	5.000
G_Sub	35,001	0.056	0.000	0.230	0.000	1.000
Proposed MRR Vetoed (period 2)	22,500	0.035	0.000	0.184	0.000	1.000
MRR Both Periods (period 1)	12,501	0.099	0.000	0.299	0.000	1.000
MRR Both Periods (period 2)	22,500	0.053	0.000	0.224	0.000	1.000
MRR Period 1 Only (period 1)	12,501	0.259	0.000	0.438	0.000	1.000
MRR Period 1 Only (period 2)	22,500	0.017	0.000	0.131	0.000	1.000
MRR Period 2 Only (period 2)	22,500	0.466	0.000	0.499	0.000	1.000

Table 2
Determinants of a (mutual) nomination
This table presents results for OLS analyses examining the determinants of a potential giver being nominated to rate the receiver (column 1) and the determinants of a mutual nomination, provided a potential giver has been nominated (column 2). The sample comprises potential nominations in period 2 (column 1) and actual nominations in period 2 (column 2). We exclude nominations in column 2 where the giver is a primary subordinate of the receiver because, by design, a mutual nomination cannot exist. Standard errors appear below the coefficients and are clustered by both receiver and employee pair. *, **, *** indicate statistical significance at the 10, 5, 1% level, respectively (two-tailed). All variables are defined in Appendix A.

		Nominee	*Mutual Nomination* (vs. One-sided)
	Pred.	(1)	(2)
Demographic Similarity	+	0.001***	0.019**
		(0.000)	(0.008)
Organizational Proximity	+	0.024***	0.144***
		(0.000)	(0.008)
Noncompetitive Promotion Prospects	+	0.000***	0.029
		(0.000)	(0.022)
Receiver FE		Yes	Yes
Giver FE		Yes	Yes
Sample		Period 2 Potential Nominations	Period 2 Actual Nominations
Observations		17,405,889	26,365
Adj. R-squared		0.009	0.235

Table 3
Supervisor involvement in rating formation process
This table presents results for two types of OLS analyses. The first two columns examine if a supervisor veto of a rater nomination varies with whether or not the nomination is mutual. The sample comprises rater nominations in period 2. The last column three examines the determinants of a supervisor creating a mutual rating request with a rater addition. The sample comprises supervisor rater additions in period 2, excluding additions where the giver is a primary subordinate of the receiver because, by design, a mutual rating request cannot exist. Standard errors appear below the coefficients and are clustered by supervisor of the receiver (in those instances where the supervisor-ID is missing, we replace the supervisor-ID variable with the same, artificial ID so as to retain the observation). *, **, *** indicate statistical significance at the 10, 5, 1% level, respectively (two-tailed). Receiver Controls are *R_Female*, *R_Native*, *R_Age 30-40*, *R_Age 40-50*, *R_Age 50+*, *R_Tenure*, and *R_Rank*. Giver Controls are *G_Female*, *G_Native*, *G_Age 30-40*, *G_Age 40-50*, *G_Age 50+*, *G_Tenure*, *G_Rank*, and *G_Sub*. All variables are defined in Appendix A.

	Supervisor Vetoed Rater Nomination		*Supervisor-Created Mutual Rating Request*
	(1)	(2)	(3)
Mutual Nomination	-0.070***	-0.073***	
	(0.008)	(0.007)	
Demographic Similarity	0.007	0.010*	0.012
	(0.005)	(0.005)	(0.024)
Organizational Proximity	-0.010*	-0.017***	0.108***
	(0.006)	(0.005)	(0.029)
Noncompetitive Promotion Prospects	-0.013	0.003	-0.029
	(0.012)	(0.011)	(0.067)
R_Nominations	0.017***		
	(0.002)		
R_Approved Nominations To Rate			0.042***
			(0.003)
R_Pct. Mutual Approved Nominations			-0.206***
			(0.043)
Receiver Controls	All	None	All
Giver Controls	All	G_Sub	All except G_Sub
R_Department FE	Yes	No	Yes
G_Department FE	Yes	No	Yes
Receiver FE	No	Yes	No
Giver FE	No	Yes	No
Sample	Period 2 Nominations		Period 2 Supervisor Rater Additions
Observations	28,197	27,976	1,620
Adj. R-squared	0.089	0.259	0.194

Table 4, Panel A
Mutual approved rating request and rating request declined
This table presents results for OLS analyses examining if a rating request decline varies with whether or not the request is a mutual approved rating request (i.e., the request will give rise to a mutual rating relationship if both sides accept the request received). The sample comprises approved rating requests (i.e., approved employee nominations, plus any supervisor additions) in period 2. Standard errors appear below the coefficients and are clustered by both receiver and employee pair. *, **, *** indicate statistical significance at the 10, 5, 1% level, respectively (two-tailed). Receiver Controls are *R_Female*, *R_Native*, *R_Age 30-40*, *R_Age 40-50*, *R_Age 50+*, *R_Tenure*, and *R_Rank*. Giver Controls are *G_Female*, *G_Native*, *G_Age 30-40*, *G_Age 40-50*, *G_Age 50+*, *G_Tenure*, *G_Rank*, and *G_Sub*. All variables are defined in Appendix A.

		Rating Request Declined	
	Pred	(1)	(2)
Mutual Approved Rating Request	-	-0.096***	-0.088***
		(0.004)	(0.004)
Demographic Similarity		0.005	0.002
		(0.003)	(0.004)
Organizational Proximity		-0.018***	-0.020***
		(0.003)	(0.004)
Noncompetitive Promotion Prospects		-0.009	0.007
		(0.008)	(0.010)
G_Requests To Rate		0.003***	
		(0.000)	
R_Requests To Be Rated		0.003***	
		(0.001)	
Receiver Controls		All	None
Giver Controls		All	G_Sub
R_Department FE		Yes	No
G_Department FE		Yes	No
Receiver FE		No	Yes
Giver FE		No	Yes
Sample		Period 2 Approved Rating Requests	
Observations		27,086	26,831
Adj. R-squared		0.061	0.140

Table 4, Panel B

Supervisor vetoes and rating request declines at the designated giver level

This table presents results for OLS analyses examining if the likelihood of a designated giver declining any of the rating requests they receive or the percentage of the rating requests they decline varies with supervisor vetoing of nominations for them to be rated. The sample comprises all designated givers (i.e., employees who received at least one rating request) in period 2. Giver Controls are *G_Female*, *G_Native*, *G_Age 30-40*, *G_Age 40-50*, *G_Age 50+*, *G_Tenure*, *and G_Rank*. Standard errors appear below the coefficients and are clustered by supervisor of the giver (in those instances where the supervisor-ID is missing, we replace the supervisor-ID variable with the same, artificial ID so as to retain the observation). *, **, *** indicate statistical significance at the 10, 5, 1% level, respectively (two-tailed). All variables are defined in Appendix A.

	Any Requests Declined	*Pct. Requests Declined*
	(1)	(2)
G_Any Supervisor Vetoes-Mutual	0.065***	
	(0.018)	
G_Any Supervisor Vetoes-One-sided	-0.019	
	(0.016)	
G_Pct. Supervisor Vetoes-Mutual		0.026**
		(0.010)
G_Pct. Supervisor Vetoes-One-sided		-0.003
		(0.008)
G_Requests To Rate	0.035***	0.004***
	(0.002)	(0.001)
G_Pct. Mutual Nominations To Rate	-0.279***	-0.073***
	(0.024)	(0.008)
Giver Controls	All	All
G_Department FE	Yes	Yes
Sample	Period 2 Designated Givers	
Observations	3,952	3,952
Adj. R-squared	0.184	0.079
p(Supervisor Vetoes Mutual = Supervisor Vetoes One-sided)	0.004	0.058

Table 5
Determinants of a mutual rating relationship
This table presents results for OLS analyses examining the determinants of a mutual rating relationship, provided the giver rated the receiver. The sample comprises final rating relationships in period 2 (columns 1 and 2) and in both periods (columns 3 and 4). We exclude instances where the giver is a primary subordinate of the receiver because, by design, a mutual rating relationship cannot exist. Standard errors appear below the coefficients and are clustered by receiver and employee pair. *, **, *** indicate statistical significance at the 10, 5, 1% level, respectively (two-tailed). Receiver Controls are *R_Female*, *R_Native*, *R_Age 30-40*, *R_Age 40-50*, *R_Age 50+*, *R_Tenure*, and *R_Rank*. Giver Controls are *G_Female*, *G_Native*, *G_Age 30-40*, *G_Age 40-50*, *G_Age 50+*, *G_Tenure*, *G_Rank*. All variables are defined in Appendix A.

	Mutual Rating Relationship (vs. One-sided)			
	(1)	(2)	(3)	(4)
Demographic Similarity	0.020**	0.021**	0.018**	0.018**
	(0.009)	(0.009)	(0.007)	(0.007)
Organizational Proximity	0.159***	0.148***	0.155***	0.146***
	(0.009)	(0.010)	(0.007)	(0.008)
Noncompetitive Promotion Prospects	0.058***	0.034	0.052***	0.049***
	(0.020)	(0.025)	(0.017)	(0.019)
Receiver Controls	All	None	All	R_Rank
Giver Controls	All	None	All	G_Rank
R_Department FE	Yes	No	Yes	Yes
G_Department FE	Yes	No	Yes	Yes
Receiver FE	No	Yes	No	Yes
Giver FE	No	Yes	No	Yes
Period FE	N/A	N/A	Yes	Yes
Sample	Period 2 Final Rating Relationships		Both Periods Final Rating Relationships	
Observations	21,255	20,840	33,341	32,981
Adj. R-squared	0.091	0.234	0.118	0.227

Table 6, Panel A
Mutual rating relationship and rating received

This table presents results for OLS analyses examining whether the rating received depends on whether the receiver and giver are in a mutual rating relationship. The sample comprises the ratings for both periods. Standard errors appear below the coefficients and are clustered by receiver and employee pair. *, **, *** indicate statistical significance at the 10, 5, 1% level, respectively (two-tailed). Receiver Controls are *R_Female*, *R_Native*, *R_Age 30-40*, *R_Age 40-50*, *R_Age 50+*, *R_Tenure*, and *R_Rank*. Giver Controls are *G_Female*, *G_Native*, *G_Age 30-40*, *G_Age 40-50*, *G_Age 50+*, *G_Tenure*, *G_Rank*, and *G_Sub*. All variables are defined in Appendix A.

	Rating Received					
	(1)	(2)	(3)	(4)	(5)	(6)
Mutual Rating Relationship (MRR)	0.210***	0.187***	0.212***	0.190***	0.206***	0.183***
	(0.008)	(0.009)	(0.008)	(0.009)	(0.011)	(0.013)
Demographic Similarity			0.047***	0.052***	0.057***	0.065***
			(0.008)	(0.009)	(0.011)	(0.012)
Organizational Proximity			-0.018**	-0.026***	-0.028**	-0.041***
			(0.009)	(0.010)	(0.012)	(0.013)
Noncompetitive Promotion Prospects			0.147***	0.112***	0.092***	0.057**
			(0.019)	(0.022)	(0.024)	(0.027)
*MRR * Demographic Similarity*					-0.020	-0.027
					(0.016)	(0.017)
*MRR * Organizational Proximity*					0.020	0.029*
					(0.016)	(0.017)
*MRR * Noncompetitive Promotion Prospects*					0.112***	0.112***
					(0.036)	(0.037)
R_Mean Rating Received	0.565***		0.554***		0.554***	
	(0.009)		(0.009)		(0.009)	
G_Mean Rating Given	0.452***		0.452***		0.452***	
	(0.009)		(0.009)		(0.009)	
Receiver Controls	All	R_Rank	All	R_Rank	All	R_Rank
Giver Controls	All	G_Rank and G_Sub	All	G_Rank and G_Sub	All	G_Rank and G_Sub
R_Department FE	Yes	Yes	Yes	Yes	Yes	Yes
G_Department FE	Yes	Yes	Yes	Yes	Yes	Yes
Receiver FE	No	Yes	No	Yes	No	Yes
Giver FE	No	Yes	No	Yes	No	Yes
Period FE	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Both Periods First-stage Ratings					
Observations	35,001	34,943	35,001	34,820	35,001	34,820
Adj. R-squared	0.262	0.394	0.264	0.396	0.264	0.396

Table 6, Panel B

Mutual rating relationship and rating received: Actual mutual rating relationship as strategic rating incentive

This table presents results for OLS analyses examining whether the rating received is higher if the receiver and giver are in an existing mutual rating relationship, as opposed to an ex-ante vetoed mutual rating relationship. The sample comprises the ratings for period 2 (we use period 2 only because the analysis requires nomination data). Standard errors appear below the coefficients and are clustered by receiver and employee pair. *, **, *** indicate statistical significance at the 10, 5, 1% level, respectively (two-tailed). Receiver Controls are *R_Female*, *R_Native*, *R_Age 30-40*, *R_Age 40-50*, *R_Age 50+*, *R_Tenure*, and *R_Rank*. Giver Controls are *G_Female*, *G_Native*, *G_Age 30-40*, *G_Age 40-50*, *G_Age 50+*, *G_Tenure*, *G_Rank*, and *G_Sub*. All variables are defined in Appendix A.

	Rating Received	
	(1)	(2)
Mutual Rating Relationship	0.239***	0.211***
	(0.010)	(0.012)
Proposed MRR Vetoed	0.134***	0.150***
	(0.023)	(0.028)
Demographic Similarity	0.047***	0.050***
	(0.010)	(0.011)
Organizational Proximity	-0.035***	-0.030**
	(0.011)	(0.012)
Noncompetitive Promotion Prospects	0.137***	0.062**
	(0.023)	(0.029)
R_Mean Rating Received	0.576***	
	(0.011)	
G_Mean Rating Given	0.482***	
	(0.011)	
Receiver Controls	All	None
Giver Controls	All	G_Sub
R_Department FE	Yes	No
G_Department FE	Yes	No
Receiver FE	No	Yes
Giver FE	No	Yes
Sample	Period 2 First-stage Ratings	
Observations	22,497	22,320
Adj. R-squared	0.272	0.435
p(Mutual Rating Relationship = Proposed MRR Vetoed)	0.000	0.028

Table 6, Panel C

Mutual rating relationship and rating received: Relationship persistence as strategic rating incentive

This table presents results for OLS analyses examining whether the mutual rating premium varies with the persistence of the mutual rating relationship. The sample comprises the ratings in period 1 (columns 1 and 2) and the ratings in period 2 (columns 3 and 4). Standard errors appear below the coefficients and are clustered by receiver and employee pair. *, **, *** indicate statistical significance at the 10, 5, 1% level, respectively (two-tailed). Receiver Controls are *R_Female*, *R_Native*, *R_Age 30-40*, *R_Age 40-50*, *R_Age 50+*, *R_Tenure*, and *R_Rank*. Giver Controls are *G_Female*, *G_Native*, *G_Age 30-40*, *G_Age 40-50*, *G_Age 50+*, *G_Tenure*, *G_Rank*, and *G_Sub*. All variables are defined in Appendix A.

	Rating Received			
	(1)	(2)	(3)	(4)
MRR Both Periods	0.239***	0.197***	0.292***	0.255***
	(0.022)	(0.028)	(0.023)	(0.025)
MRR Period 1 Only	0.164***	0.094***	0.092***	0.087**
	(0.016)	(0.019)	(0.034)	(0.037)
MRR Period 2 Only			0.225***	0.194***
			(0.010)	(0.012)
Demographic Similarity	0.047***	0.038**	0.046***	0.049***
	(0.014)	(0.017)	(0.010)	(0.011)
Organizational Proximity	0.011	0.032*	-0.036***	-0.030**
	(0.015)	(0.019)	(0.011)	(0.012)
Noncompetitive Promotion Prospects	0.158***	0.056	0.137***	0.064**
	(0.033)	(0.044)	(0.023)	(0.029)
R_Mean Rating Received	0.519***		0.576***	
	(0.014)		(0.011)	
G_Mean Rating Given	0.404***		0.482***	
	(0.014)		(0.011)	
Receiver Controls	All	None	All	None
Giver Controls	All	G_Sub	All	G_Sub
R_Department FE	Yes	No	Yes	No
G_Department FE	Yes	No	Yes	No
Receiver FE	No	Yes	No	Yes
Giver FE	No	Yes	No	Yes
Sample	Period 1 First-stage Ratings		Period 2 First-stage Ratings	
Observations	12,501	12,098	22,497	22,320
Adj. R-squared	0.253	0.440	0.272	0.435
p(MRR Both Periods = MRR Period 1 Only)	0.002	0.000	0.000	0.000
p(MRR Both Periods = MRR Period 2 Only)			0.003	0.009
p(MRR Period 1 Only = MRR Period 2 Only)			0.000	0.005

Table 7
Final supervisor rating: Weights on mutual ratings vs. one-sided ratings
This table presents results for OLS analyses examining the implicit weights placed by supervisors on mutual vs. one-sided ratings, in the full sample and in different subsamples. The sample comprises supervisor ratings in period 2. Standard errors appear below the coefficients and are clustered by supervisor. *, **, *** indicate statistical significance at the 10, 5, 1% level, respectively (two-tailed). Receiver Controls are *R_Female, R_Native, R_Age 30-40, R_Age 40-50, R_Age 50+, R_Tenure*, and *R_Rank*. All variables are defined in Appendix A.

	Supervisor Rating			
	(1)	(2)	(3)	(4)
R_Weighted Mean Mutual Rating	0.415***	0.419***	0.335***	0.518***
	(0.018)	(0.020)	(0.028)	(0.031)
R_Weighted Mean One-sided Rating	0.399***	0.405***	0.302***	0.531***
	(0.020)	(0.022)	(0.031)	(0.033)
R_Pct. Nominations Vetoed	-0.050	-0.050	-0.023	-0.013
	(0.040)	(0.059)	(0.077)	(0.105)
R_All Raters Supervisor Additions	-0.032	-0.044	-0.006	-0.030
	(0.038)	(0.041)	(0.054)	(0.099)
Receiver Controls	All	All	All	All
R_Department FE	Yes	Yes	Yes	Yes
Supervisor FE	No	Yes	Yes	Yes
Sample	Period 2 All Supervisor Ratings		Period 2 Receiver has <= 5 First-stage Ratings	Period 2 Receiver has >5 First-stage Ratings
Observations	3,332	3,249	1,504	1,539
Adj. R-squared	0.235	0.253	0.201	0.291
p(*R_Weighted Mean Mutual Rating* = *R_Weighted Mean One-sided Rating*)	0.055	0.113	0.006	0.437

Appendix A
Variable definitions

Rating and Relationship Variables	
Any Requests Declined	Indicator variable equal to 1 if a designated giver declines at least one rating request, and 0 otherwise.
G_Any Supervisor Vetoes-Mutual	Indicator variable equal to 1 if a supervisor vetoed any mutual nominations for the designated giver to be rated, and 0 otherwise.
G_Any Supervisor Vetoes-One-sided	Indicator variable equal to 1 if a supervisor vetoed any one-sided nominations for the designated giver to be rated, and 0 otherwise.
G_Pct. Mutual Nominations To Rate	Percentage of the approved nominations for the designated giver to provide ratings that are mutual.
G_Pct. Supervisor Vetoes-Mutual	Percentage of the mutual nominations for the designated giver to be rated that were vetoed by a supervisor (equals 0 if no mutual nominations were vetoed or if there were no mutual nominations).
G_Pct. Supervisor Vetoes-One-sided	Percentage of the one-sided nominations for the designated giver to be rated that were vetoed by a supervisor (equals 0 if no one-sided nominations were vetoed or if there were no one-sided nominations).
MRR Both Periods	Indicator variable equal to 1 if, for the receiver-giver pair, *Mutual Rating Relationship* is equal to 1 for period 1 and period 2; 0 otherwise.
MRR Period 1 Only	Indicator variable equal to 1 if, for the receiver-giver pair, *Mutual Rating Relationship* is equal to 1 in period 1 but not period 2; 0 otherwise.
MRR Period 2 Only	Indicator variable equal to 1 if, for the receiver-giver pair, *Mutual Rating Relationship* is equal to 1 in period 2 but not period 1; 0 otherwise.
Mutual Approved Rating Request	Indicator variable equal to 1 if the designated giver receives an approved request to rate the receiver (either a supervisor-approved nomination or a supervisor addition) *and* the receiver receives an approved request to rate the giver, and equal to 0 if the designated giver receives an approved request to rate the receiver but the receiver *does not* receive an approved request to rate the giver.
Mutual Nomination	Indicator variable equal to 1 if the receiver is nominated to rate the potential giver *and* the potential giver is nominated to rate the receiver, and equal to 0 if the potential giver is nominated to rate the receiver but the receiver *is not* nominated to rate the potential giver.
Mutual Rating Relationship	Indicator variable equal to 1 if the receiver received a rating from the giver *and* provided the giver with a rating, equal to 0 if the receiver received a rating from the giver and *did not* provide the giver with a rating (i.e., it is a one-sided rating relationship).
Nominee	Indicator equal to 1 if the receiver nominates the potential giver to provide a rating, and equals 0 if the receiver does not nominate the potential giver.
Pct. Requests Declined	Percentage of the number of ratings requests declined by the designated giver.
Proposed MRR Vetoed	Indicator variable equal to 1 if the receiver received a rating from the giver in the current period and the receiver was nominated to rate the giver but the nomination was vetoed by the supervisor (i.e., there would have been a mutual rating relationship if not for a supervisor veto); 0 otherwise.
Rating Received	Overall rating the receiver received from the giver, on a scale from 1 (lowest) to 5 (highest).
Rating Request Declined	Indicator variable equal to 1 if the designated giver declined to rate the receiver and 0 if the designated giver did not decline the rating request. Rating requests that are not declined may have the status “finished”, “started”, or “pending/waiting”.

Supervisor-Created Mutual Rating Request	Indicator variable equal to 1 if the supervisor rater addition resulted in a mutual rating request and 0 if the supervisor rater addition resulted in a one-sided rating request.
Supervisor Rating	Rating the employee (receiver) received from their supervisor; 1 = "needs improvement," 2 = "good performer," and 3 = "top performer."
Supervisor Vetoed Rater Nomination	Indicator variable equal to 1 if the supervisor vetoed the nomination for the receiver to be rated by the potential giver and 0 if the supervisor approved the nomination.
R_All Raters Supervisor Additions	Indicator variable equal to 1 if all of the receiver's raters are the result of supervisor additions (i.e., none from employee nominations); 0 otherwise.
R_Pct. Mutual Approved Nominations	Percentage of the receiver's approved nominations to be rated that are mutual (i.e., the receiver is also approved to rate the giver).
R_Pct. Nominations Vetoed	Percentage of the nominations for the receiver to be rated that were vetoed by the supervisor.
R_Weighted Mean Mutual Rating	Average of all mutual ratings the receiver received in the first stage (i.e., ratings where *Mutual Rating Relationship* is equal to 1) multiplied by the percentage of ratings that are mutual. This variable equals 0 if the receiver has no mutual ratings.
R_Weighted Mean One-sided Rating	Average of all one-sided ratings the receiver received in the first stage (i.e., only ratings where *Mutual Rating Relationship* is equal to 0) multiplied by the percentage of ratings that are one-sided. This variable equals 0 if the receiver has no one-sided ratings.
Demographic Similarity Variables	
Demographic Similarity	Indicator variable equal to 1 if at least three of *Same Gender*, *Same Age, Same Nationality*, and *Same Tenure* are equal to 1; 0 otherwise.
Same Age	Indicator variable equal to 1 if the receiver and giver belong to the same age category (e.g., 30 to 40); 0 otherwise (we use age category as the data we received indicated only an employee's age category).
Same Gender	Indicator variable equal to 1 if the receiver and giver both identify as female or both identify as male; 0 otherwise.
Same Nationality	Indicator variable equal to 1 if the receiver and giver are both natives of the country where the company is headquartered or the receiver and giver are both international; 0 otherwise.
Same Tenure	Indicator variable equal to 1 if the receiver and giver joined the company within six months of each other; 0 otherwise.
Organizational Proximity Variables	
Organizational Proximity	Indicator variable equal to 1 if at least two of *Same Supervisor*, *Same Department*, and *Same Rank* are equal to 1; 0 otherwise.
Same Department	Indicator variable equal to 1 if the receiver and giver belong to the same department; 0 otherwise.
Same Rank	Indicator variable equal to 1 if the receiver and giver are at the same hierarchical rank and there is no supervisor/subordinate relationship between the two; 0 otherwise.
Same Supervisor	Indicator variable equal to 1 if the receiver and giver have the same supervisor; 0 otherwise.
Rater Cooperation Incentives	
Noncompetitive Promotion Prospects	Indicator variable equal to 1 if the receiver and giver are both considered "ready for promotion" but are in different departments and are at different ranks; 0 otherwise.
Receiver Controls	
R_Age <=30	Indicator variable equal to 1 if the receiver is aged 30 or below; 0 otherwise.
R_Age 30-40	Indicator variable equal to 1 if the receiver is aged between 30 and 40; 0 otherwise.

R_Age 40-50	Indicator variable equal to 1 if the receiver is aged between 40 and 50; 0 otherwise.
R_Age 50+	Indicator variable equal to 1 if the receiver is aged 50 or above; 0 otherwise.
R_Approved Nominations to Rate	Number of approved nominations for the receiver to provide ratings.
R_Female	Indicator variable equal to 1 if the receiver identifies as female; 0 otherwise.
R_Mean Rating Received	The average rating *received* by the receiver for the period, excluding the focal rating observation.
R_Native	Indicator variable equal to 1 if the receiver is a native of the country where the company is headquartered; 0 otherwise.
R_Rank	Hierarchical rank of the receiver, on a scale from 0 through 5, where a higher value corresponds to a higher rank.
R_Requests to be Rated	Number of requests (approved nominations and supervisor additions) sent to designated givers to rate the receiver.
R_Tenure	Tenure of the receiver with the company in years.
R_Nominations	Number of nominations for the receiver to be rated (before supervisor vetoes).
Giver Controls	
G_Age <=30	Indicator variable equal to 1 if the giver is aged 30 or below; 0 otherwise.
G_Age 30-40	Indicator variable equal to 1 if the giver is aged between 30 and 40; 0 otherwise.
G_Age 40-50	Indicator variable equal to 1 if the giver is aged between 40 and 50; 0 otherwise.
G_Age 50+	Indicator variable equal to 1 if the giver is aged 50 or above; 0 otherwise.
G_Female	Indicator variable equal to 1 if the giver identifies as female; 0 otherwise.
G_Mean Rating Given	The average rating *given* by the giver for the period, excluding the focal rating observation.
G_Native	Indicator variable equal to 1 if the giver is a native of the country where the company is headquartered; 0 otherwise.
G_Rank	Hierarchical rank of the giver, on a scale from 0 through 5, where a higher value corresponds to a higher rank.
G_Requests to Rate	Number of requests (approved nominations and supervisor additions) the designated giver received to rate receivers.
G_Sub	Indicator variable equal to 1 if the giver is a primary subordinate of the receiver (that is, the receiver is the giver's primary supervisor); 0 otherwise.
G_Tenure	Tenure of the giver with the company in years.

Appendix B
Samples and inclusion criteria

Sample A.1
Primary sample at the potential nomination level (only available for period 2): 17,405,889 observations

Inclusion criteria

- Receiver and giver employee-ID numbers populated
- Complete demographic, department and rank data for the receiver and for the giver
- Employee-ID number for the receiver's and/or the giver's supervisor is populated, except when *Organizational Proximity* is not affected
- The giver is not the receiver's primary supervisor (the primary supervisor evaluates the receiver in the second stage of the evaluation process)

Sample A.2
Primary sample at the nomination level (only available for period 2): 28,199 observations

Inclusion criteria

- Receiver and giver employee-ID numbers populated
- Complete demographic, department and rank data for the receiver and for the giver
- Employee-ID number for the giver's supervisor is populated, except when *Organizational Proximity* is not affected
- The giver is not the receiver's primary supervisor (the primary supervisor evaluates the receiver in the second stage of the evaluation process, so we exclude any primary supervisor nominations in the first stage)

Sample A.3
Primary sample at the rater addition level (only available for period 2): 1,629 observations

Inclusion criteria

- Receiver and giver employee-ID numbers populated
- Complete demographic, department and rank data for the receiver and for the giver
- Employee-ID number for the giver's supervisor is populated, except when *Organizational Proximity* is not affected
- The giver is not the receiver's primary supervisor (the primary supervisor evaluates the receiver in the second stage of the evaluation process, so we exclude any primary supervisor additions in the first stage)
- The giver is not a primary subordinate of the receiver (as there cannot be a mutual rating relationship by definition)
- The number of approved nominations for the receiver to provide a rating for others (where giver employee-ID number is populated) can be calculated and is greater than zero

Sample A.4
Primary sample at the rating request level (only available for period 2): 27,089 observations

Inclusion criteria

- Receiver and giver employee-ID numbers populated
- Complete demographic, department and rank data for the receiver and for the giver
- Employee-ID number for the receiver's and/or the giver's supervisor is populated, except when *Organizational Proximity* is not affected
- The number of the requests sent for the receiver to be rated can be calculated
- The number of the requests received by the giver to provide a rating can be calculated
- The giver is not the receiver's primary supervisor (the primary supervisor evaluates the receiver in the second stage of the evaluation process, so we exclude any primary supervisor requests in the first stage)

Sample A.5
Primary sample at the individual rating level: 35,001 observations

Inclusion criteria

- Receiver and giver employee-ID numbers populated
- Complete demographic, department, rank, and promotion readiness data for the receiver and for the giver
- Employee-ID number for the receiver's and/or the giver's supervisor is populated, except when *Organizational Proximity* is not affected
- Rating has the status "finished" (i.e., started but not completed ratings are excluded)
- *R_Mean Rating Received* can be calculated (requires that the receiver received at least one other rating during the period)
- *G_Mean Rating Given* can be calculated (requires that the giver gave at least one other rating during the period)
- The giver is not the receiver's primary supervisor (the primary supervisor evaluates the receiver in the second stage of the evaluation process, so we exclude any primary supervisor first-stage ratings)

Sample A.6
Primary sample at the supervisor-rating level: 3,334 observations

Inclusion criteria

- Is a period 2 observation
- Receiver employee-ID number populated
- Underlying first-stage ratings have the giver employee-ID number populated (needed to determine mutual ratings)
- Primary supervisor employee-ID number populated
- Complete demographic, department, and rank data for the receiver

Appendix C
Period 2 rating sample reconciliation

	Useable observations
Rater nominations from employees	28,199
Less: Supervisor vetoes of employee nominations	3,184
Plus: Supervisor rater requests	2,074
Total rater requests	27,089
Less: Requests declined	1,796
Less: Requests pending/waiting	1,835
Less: Ratings started but not completed	780
Total completed ratings	22,678
Less: Individuals for whom we can't calculate the mean rating received variables	178
Total ratings sample	22,500