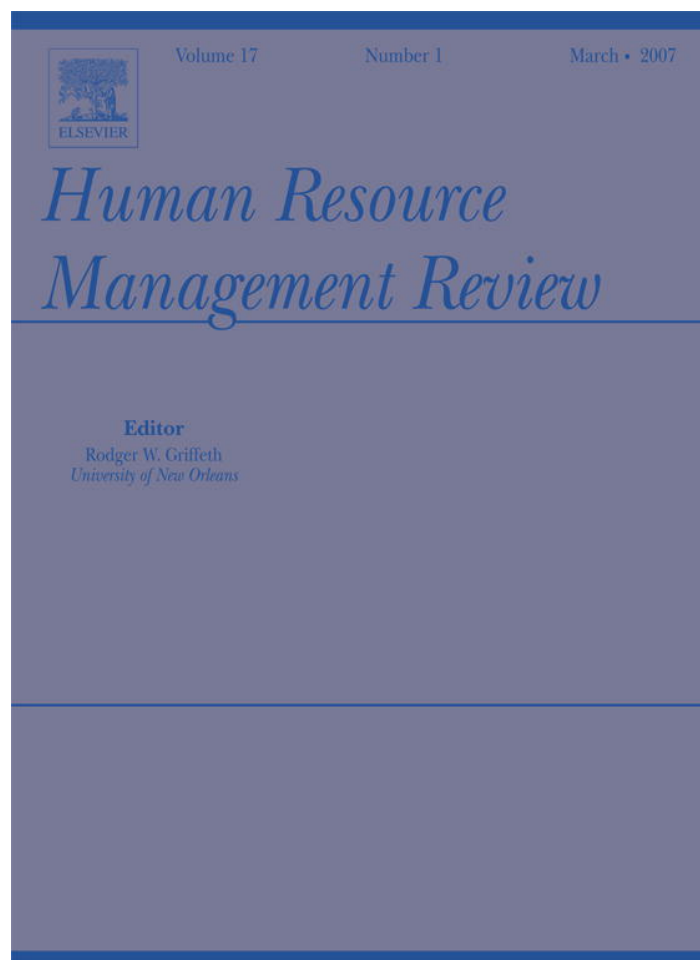


Provided for non-commercial research and educational use only.
Not for reproduction or distribution or commercial use.



This article was originally published in a journal published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues that you know, and providing a copy to your institution's administrator.

All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

<http://www.elsevier.com/locate/permissionusematerial>

Competitive appraising: A social dilemma perspective on the conditions in which multi-round peer evaluation may result in counter-productive team dynamics[☆]

Peter A. Bamberger*

Davidson Faculty of Industrial Engineering and Management, Technion — Israel Institute of Technology, Technion City, Haifa 32000, Israel

Abstract

Drawing from the literature on social dilemmas, I develop a process model identifying the conditions under which peer evaluation, when implemented over multiple rounds, might be expected to lead to increasingly counter-productive team dynamics over time. Noting that peer evaluation encompasses a set of behaviors just like those being assessed, I generate a number of hypotheses grounded on the proposition that shared perceptions regarding three system dimensions – rating purpose, rater accountability and evaluation regularity – influence the implicit incentive structures governing team members' rating behaviors, and, as such, the prevalence of intentional rating distortion in teams. The prevalence of intentional distortion, in turn, is posited to influence the degree to which direct and indirect rating-distortion experiences pervade the team environment — what I refer to as ambient peer-rating distortion. Finally, the level of team ambient peer-rating distortion is posited to influence the prevalence of actual intentional distortion in subsequent rounds of peer evaluation, and to attenuate and suppress any beneficial effect of peer evaluation on cooperation-related team processes.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Peer assessment; Intentional distortion in performance appraisal; Social dilemmas; Multisource feedback

Over the past decade, managers have increasingly sought to organize work on the basis of interdependent teams (Cohen & Bailey, 1997). As described by Wageman (1995) such teams are typically structured around two different frameworks: task or process interdependence (in which jobs are rotated, resources are shared, and tasks are often performed collectively, as in advanced manufacturing cells), and outcome interdependence (in which members work toward a collective goal but are responsible for separate tasks, as is common in software development). In the context of such team-based work frameworks, peer evaluations can provide valuable information regarding, for example, the extent to which particular team members cooperate with their peers or contribute to some collective objective. For this reason, peer evaluation has become particularly attractive to organizations adopting such frameworks (Levy & Williams, 2004; Miller & Cardy, 2000).

[☆] An earlier version of this manuscript was presented at the 2004 Annual Meeting of the Academy of Management, New Orleans, LA. The author is grateful to the following colleagues for their comments on an earlier draft of this paper: Samuel B. Bacharach, Miriam Erez, Avraham Kluger, David Waldman and Dov Zohar.

* Tel.: +972 4 829 4510; fax: +972 4 829 5688.

E-mail address: peterb@tx.technion.ac.il.

Despite the increasing popularity of peer evaluation, few studies have addressed the possible implications of peer evaluation on team-related outcomes (Drexler, Beehr, & Stetz, 2001; Levy & Williams, 2004), and even fewer have considered the nature of these team-level effects over multiple rounds of evaluation (for an exception, see Bamberger, Erev, Kimmel, & Oref-Chen, 2005). Rather, research on peer evaluation has generally been conducted at the individual level of analysis, and has largely focused on individual-level outcomes such as post-feedback task performance, attitudes and perceptions (for a review, see London & Smither, 1995). Moreover, those few studies examining the impact of peer evaluation on group- or team-related outcomes have generated largely mixed results (Levy & Williams, 2004: 897). For example, although a number of researchers have found peer assessment to improve relationships among team members and their perceptions of communication, cooperation and performance (Druskat & Wolff, 1999; Erez, Lepine, & Elms, 2002; Hazucha, Hezlett, & Schneider, 1993), others have found it to generally harm team relationships and impair group processes. Specifically, a number of studies suggest that while the effectiveness of peer evaluation is contingent upon participants' willingness to differentiate among their coworkers, many workers hesitate to do so, perhaps out of fear that they will harm the team's social climate or their relations with their teammates (Drexler et al., 2001; Kane & Lawler, 1978; Mitchell & Liden, 1982; Naiper & Latham, 1986). To the extent that participants nevertheless do cooperate, their appraisals may be subject to rater biases (Saavedra & Kwun, 1993) that can affect members' perceptions about the group and group members' team satisfaction and cohesiveness (DeNisi, Randolph, & Blencoe, 1983). Others note that peer assessments might have negative effects on group morale (Cederblom & Lounsbury, 1980) and member relations (Kane & Lawler, 1978; Love, 1981).

Viewing the work team as the primary unit of analysis, my objective in this paper is therefore to generate a theory explaining the variable impact that multi-round peer evaluation may have on such team processes over time. I ground this theory on the notion that peer evaluation encompasses a set of behaviors just like those being assessed. And like the behaviors being evaluated, these evaluation behaviors are also likely to be influenced by implicit, dynamic and often countervailing incentive structures (Dawes, 1980: 174). In this sense, I approach peer evaluation from a social dilemma perspective with a particular focus on intentional rater distortion, defined by Kozlowski, Chao and Morrison (1998: 164) as relating to participants' *willingness* to accurately report judgments of their coworkers' performance, rather than their *ability* to make such judgments in the first place.

A social dilemma is a paradox generated by the dual utility functions often confronted by individuals who share responsibility for some common end-result or outcome. Each individual realizes that this shared goal is unlikely to be achieved if members engage in opportunistic behavior (i.e., "free ride" off the efforts of others). At the same time, each also realizes that his or her own individual utility (share of collective or group benefit minus his or her cost) is maximized by doing precisely that (Dawes, 1980). In work organizations, individuals are often torn between such logics of collective rationality (i.e., recognition that everyone benefits if all cooperate) and individual rationality (the idea that one's own payoff is likely to rise the more one relies on others to ensure that collective objectives are met) (Albanese & Van Fleet, 1985). Social dilemmas force members of a team to choose between a collectively rational solution such as task-sharing, and an individually rational solution such as social loafing. Likewise, social dilemmas are manifest when peer raters must decide – for example – whether to intentionally distort their assessments downward so as to enhance their own relative standing (individual rationality), or to offer unbiased appraisals with the aim of enhancing team or organizational efficiencies (collective rationality).

Considering the implications of such social dilemmas on team-level processes, I suggest that *in theory*, peer evaluation should reduce the utility of social loafing to individual team members (by shifting incentive structures relating to team-oriented behaviors), and thereby generate positive team dynamics. However, *when conducted over multiple rounds*, I argue that, at least under certain conditions, these positive effects may become attenuated and suppressed. Adopting a broader time-frame and considering peer evaluation when conducted over multiple rounds, I argue that any longer-term, positive impact of peer evaluation on team processes will be conditional upon members' shared perceptions of the particular peer evaluation system being implemented. Depending on these perceived system characteristics, the team environment may, over *multiple rounds*, be increasingly pervaded by a sense that peer-ratings are intentionally and downwardly biased. Such ambient evaluation-related stimuli may not only shift incentive structures underlying evaluation behaviors (thereby increasing the utility of intentional downward distortion in subsequent rounds), but may also attenuate and suppress any positive, direct effects of peer evaluation on team processes.

By explicating this model, I seek to contribute to the literature on peer evaluation in two main ways. First, I seek to generate a theory regarding a critical, yet under-researched aspect of peer evaluation: namely, how multi-round peer

evaluation may influence key team processes, and the conditions under which positive team-related outcomes may be attenuated and suppressed. Regardless of the psychometric properties of peer evaluation and its validity as an employee development or selection tool, as recognized by Levy and Williams (2004) in their review of the peer evaluation literature, there is substantial practical value to gaining a better understanding of the broader social consequences of such a managerial tool. Second, by adopting a process-based, longitudinal mode of analysis (Mohr, 1982), I seek to provide a dynamic framework for examining such effects, one focusing on a likely “chain of events.” As such, while still concerned with explaining the variance in critical team processes, I place an emphasis on the causal ordering of events in an attempt to show how the effects of peer evaluation may unfold over time (Mohr, 1982: 44). In this way, I seek to respond to Levy and Williams’ (2004: 895) call for scholars to take peer evaluation research “in a different direction — focusing on the peer assessment process rather than solely on the more traditional measurement of team effectiveness.”

I begin my analysis by laying out the assumptions underlying my theory. Next, I begin to construct my process model of the effects of peer evaluation on team processes by reviewing the logic underlying a beneficial effect perspective of peer evaluation. I then turn to identifying the conditions likely to govern such a relationship, describing how, over multiple rounds, team members’ collective perceptions of the peer evaluation context may generate increasingly counter-productive team dynamics, with the potential to attenuate and suppress any such beneficial consequences.

1. Assumptions

Underlying this model are a number of assumptions. First, I assume that cooperation-related team processes serve as a primary criterion by which organizations should assess the team-related consequences of peer evaluation. Such an assumption is reasonable, given that peer evaluation systems are often put in place specifically to assess and influence these parameters (Bracken, Timmerck, & Church, 2001), and given that cooperative member behavior has long been assumed to be critical to effective group performance (for reviews, see Campion, Medsker, & Higgs, 1993). Recent studies of team effectiveness (Cohen & Bailey, 1997; Hackman, 1987; Murphy & Cleveland, 1995) identify a number of cooperation-related processes – namely task-sharing, knowledge-sharing, and helping – as among those likely to have the most critical effects on overall team effectiveness. Consequently, consistent with a number of recent peer evaluation studies (Druskat & Wolff, 1999; Erez et al., 2002), I adopt these three cooperation-related team processes as the primary criterion variables in the model I generate below.

Second, I assume that peer evaluation is distinct from other forms of multi-source feedback in that peer raters may perceive themselves to be in direct competition with ratees for valued resources, the distribution of which they often believe is influenced by evaluation scores (Greguras, Robie, Schleicher, & Goff, 2003). Consequently, while for supervisors and subordinate raters there may be a substantial pecuniary cost for providing negative feedback (Baron & Kreps, 1999), for peer raters there may be a substantial pecuniary cost for offering *positive* feedback. Further increasing the pecuniary costs of positive feedback to peer raters, I assume that many of these valued resources are often distributed or perceived to be distributed on the basis of a zero-sum game. In other words, participants in a peer evaluation system may believe that any increase in the rewards (e.g., compensation) distributed to a ratee on the basis of an evaluation score will reduce the pool of rewards available for distribution to the rater. The upshot is that while intentional distortion in traditional supervisory evaluation tends to result in overly lenient assessments (Jawahar & Williams, 1997), I assume that in many peer evaluation contexts, the opposite is the case: rater pecuniary costs and incentive structures mean that over multiple rounds, overly harsh assessments may become just as prevalent as overly lenient ones, if not more so. Moreover, while a tendency towards leniency may diminish the psychometric and social value of peer evaluation, such a tendency is unlikely to induce the kind of conflict-laden, retaliatory behaviors that might be expected as a result of intentional harshness among even a small number of peer raters in a team. Consequently, while not denying the potential for intentional leniency in peer evaluation, I focus my attention on intentional harshness as the rating tendency likely to have the most salient team-process implications.

Finally, I assume that while *actual* evaluation system characteristics and *actual* distortion in peer-ratings play an important role in determining team-level peer evaluation outcomes, it is the employees’ *shared perceptions* of the nature of the evaluation system, as well as their collective sense that intentional bias pervades the team environment (what I refer to later as “ambient peer-rating distortion”), that are likely to have the most profound impact on subsequent team member task and rating behaviors (Schneider, 2000). In this sense, while my hypotheses are structured around the

team as the primary unit of analysis, I implicitly view the social dilemma in peer evaluation as a multi-level phenomenon in which individual evaluation behaviors create ambient phenomena at the team-level which, in turn, influence the subsequent peer evaluation behaviors of team members.

2. Peer evaluation and team processes: beneficial effects

Despite the inconsistencies in the research literature regarding the impact of peer evaluation on team-member relations and cooperation-related team processes, theory suggests two main reasons why peer evaluation – at least when initially implemented in an interdependent team – may be expected to improve team dynamics. Specifically, such positive outcomes are likely to stem from (1) the unique information that peers can offer one another, and (2) enhanced team member monitoring and sanctioning capabilities.

A number of studies suggest that a key advantage of peer evaluation is that it offers high-quality, supplementary performance-related feedback that may be more acceptable and meaningful to team members than supervisory feedback (Murphy & Cleveland, 1991). According to Fedor and Bettenhausen (1989:182), peer evaluation “enhances both the accuracy of performance ratings and the quality of the feedback.” Because supervisors often lack the close contact with their subordinates enjoyed by peers, they may be unable to accurately and consistently observe cooperative behaviors critical to team effectiveness, like task-sharing, knowledge-sharing, and help-seeking and -giving (Murphy & Cleveland, 1995: 140). Indeed, particularly in the context of semi-autonomous teams, a far greater emphasis is placed on supervisory monitoring of team *outcomes* than of team *processes* (Erez et al., 2002). Moreover, when supervisors (as opposed to peers) are responsible for monitoring team processes, employees may try to make a good impression by accentuating certain valued but largely atypical behaviors, while avoiding other, more typical behaviors – like open displays of help-seeking – that they fear supervisors might attribute to a lack of competence (Higgins & Kram, 2001:276). Peers, in contrast, are in a unique position to observe and evaluate help-seeking behaviors, as they are likely to be the ones approached for help. What is more, the performance-related insights provided by peers may allow individuals to develop new strategies for enhancing the long-term value of help-seeking and other cooperation-related behaviors. For example, peer feedback may encourage peers to shift from a dependent mode of help-seeking (requiring the peer help-giver to actually provide the solution to some problem) to a more autonomous mode (in which the help-giver provides the tools by which others can identify the solution on their own) (Nadler, 1998). In short, team members – as opposed to supervisors – may be uniquely positioned to observe, evaluate and enhance those behaviors that are often at the core of effective team functioning.

Additionally, the literature on social dilemmas (Albanese & Van Fleet, 1985; Kerr & Brunn, 1981) suggests that the enhanced degree of monitoring and sanctioning inherent in peer evaluation may be associated with positive cooperation-related team processes. Specifically, peer evaluation may provide an attractive strategy by which to counter social loafing and encourage task- and knowledge-sharing as well as peer help-giving, in that it effectively changes the individual incentive system. Just as recommended by Albanese and Van Fleet (1985: 253), peer evaluation “builds a private good that is contingent on the provision of the group’s public good into the group member’s incentive system.” That is, it better aligns individual interests with those of the collective by making the individual’s organizational future more contingent upon his or her contribution to the group, and in that sense provides a platform for enhanced team-level outcomes.

Furthermore, by increasing the probability that team members will be caught and sanctioned for free-riding, peer evaluation may alter members’ perceptions of the costs versus benefits of social loafing. Rather than relying upon tight supervision to monitor individual effort and enhance task noticeability (Kerr & Brunn, 1981), peer evaluation diffuses the monitoring role equitably among all group members, and ensures that this monitoring is as continuous as the group activity itself. As soon as group members recognize that they, like their coworkers, are being continuously monitored by others able to impose sanctions (even if these are only normative, in the form of negative developmental feedback), they are less likely to adopt behaviors which might be seen as free-riding. Conversely, the ability to impose sanctions on free-riding coworkers makes team members less likely to resort to free-riding themselves when they feel they are being taken advantage of (i.e., are being a “sucker”). In this sense, peer assessment offers teams a means by which to break the potentially self-reinforcing cycle of non-cooperation and social loafing that may often serve as an occupational hazard of team environments.

This perspective thus suggests that in theory, peer evaluation in work teams has the potential to yield a number of positive team-level consequences, at least initially (see Fig. 1). First, when group members know that they will assess

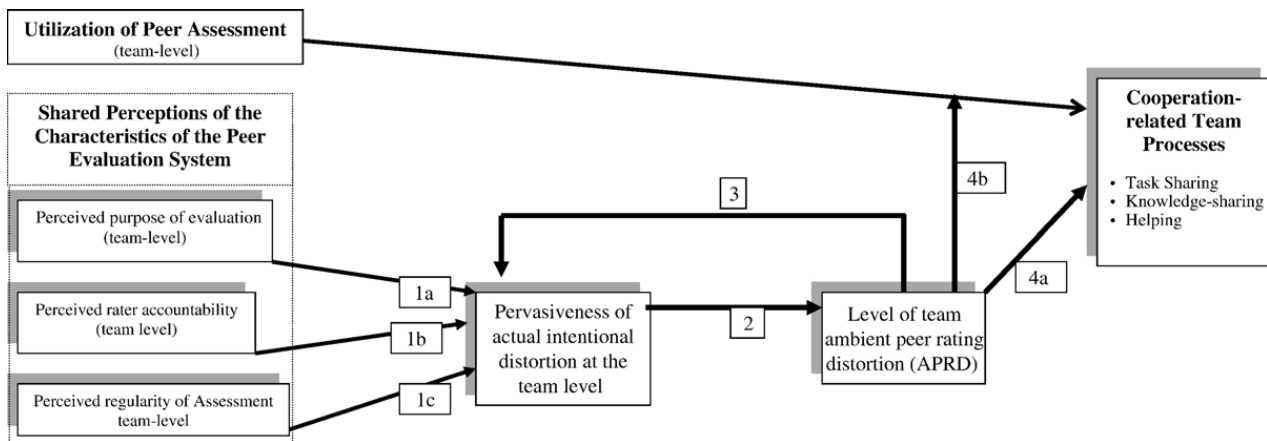


Fig. 1. A team-level, process model of the impact of peer evaluation on cooperation-related team processes.

and be assessed by their peers, the incentive to “free-ride” should be reduced — along with the incentive for others to adopt similar behavior in order to avoid becoming a “sucker”. Consequently, relative to teams in which members’ performance is evaluated strictly on the basis of supervisory appraisals, in teams in which members’ performance is also assessed on the basis of peer evaluations, the overall level of task-sharing should be higher.

Second, when team members know that their peers have the ability to reward cooperation and punish lack of cooperation, I would expect peers to develop norms supportive of knowledge-sharing and autonomous (as opposed to dependent) help-seeking and -giving. Recognizing that peer evaluation provides coworkers with an inherent interest to help one another (as a means by which to enhance one’s own peer-ratings), over time, individuals employed in such units are likely to feel more at ease in requesting the assistance of their peers. At the same time, they are likely to recognize that excessive help-seeking may generate negative peer evaluations. For this reason they will limit help-seeking to more critical issues, and will seek that kind of help which will allow them to solve similar problems in the future more independently — namely, autonomous help (Nadler, 1998). Whereas in traditionally evaluated teams, members might hesitate to provide others with proprietary knowledge or teach them skills that may serve as a basis of power (Fisher, Goff, Nadler, & Chinsky, 1988), in teams with peer evaluation such incentive frameworks are likely to be counterbalanced by the recognition that individual task performance is but one dimension of total performance — the other dimension being how well the individual helps coworkers develop and achieve their maximum potential.

Taken as a whole, the discussion above suggests that peer evaluation, at least when initially implemented, will have largely beneficial implications with regard to cooperation-related team processes. Indeed, the results of a number of recent field studies examining the social consequences of a single round of peer evaluation (e.g., Druskat & Wolff, 1999; Erez et al., 2002) are consistent with such a notion. However, as noted in the introduction, other studies (e.g., DeNisi et al., 1983) suggest that while peer evaluation may not have an immediate adverse impact on cooperation-related team processes, it may still generate negative member perceptions regarding the group, and in particular, its members’ evaluation-related tendencies. Depending on their level of pervasiveness within the group, such perceptions could conceivably have a delayed but still adverse impact on cooperation-related team processes. Consequently, in the sections below, I pay particular attention to such perceptions, focusing on the conditions likely to result in their emergence, and the way in which, over multiple rounds of peer evaluation, such perceptions might increasingly attenuate or suppress the otherwise beneficial effects of peer evaluation on cooperation-related team processes.

3. System characteristics, intentional distortion and team processes

As noted earlier, the theory (which I explicate in more detail below) examines the emergence and impact of such potentially counter-productive evaluation-related perceptions from a social dilemma perspective. Drawing from such a perspective, I argue that shared perceptions regarding the nature of this system may encourage team members to intentionally distort their evaluations of one another in a downward direction. I further posit that members’ sense that they themselves have been victims of intentional distortion (direct exposure), or awareness of other team members who may feel this way (indirect exposure), may generate a team context pervaded by a sense that peer-ratings are

downwardly biased — what I refer to as an ambient peer-rating distortion (APRD). In such team contexts, I argue that two team-level outcomes may be expected: first, the emergence of an implicit pay-off structure encouraging increasingly severe intentional distortion from one round of peer evaluation to the next; and second, the attenuation and suppression of any positive association between peer evaluation and cooperation-related team processes from round to round.

3.1. Shared perceptions, incentive structures and intentional distortion

As noted above, most of the research on performance appraisal has focused on the psychometric quality and accuracy of various appraisal techniques and, more recently, on rater cognition. Underlying most of this research is the assumption that while bias and rating inaccuracy may result in significant rating distortions, these distortions are largely unintentional. However, [Bernardin and Villanova \(1986\)](#) suggest that deliberate rater distortion may be more prevalent than unintentional bias. Similarly, [Longenecker and Ludwig \(1990\)](#) report that more than 70% of managers in their study admitted to having inflated or deflated their evaluations of subordinates in order to either protect a colleague or to shock an adversary or poor performer. Other studies suggest that the organizational and performance management context may influence rating accuracy not only by determining raters' *ability* to make accurate judgments, but also by shaping their *willingness* to do so ([Kozlowski et al., 1998](#); [Murphy & Cleveland, 1995](#)). These findings suggest that we examine performance evaluation, regardless of rating source, as goal-directed behavior, with contextual factors playing an important role in determining the incentive structures motivating more or less desirable rating behaviors ([Levy & Williams, 2004](#); [Murphy & Cleveland, 1991](#)).

In the discussion below, I attempt to shed some light on how shared perceptions regarding three critical dimensions of peer assessment systems – *evaluation purpose*, *rater accountability* and *the regularity of evaluation* ([DeNisi & Kluger, 2000](#); [Levy & Williams, 2004](#)) – may encourage intentional downward distortion on the part of peer raters. Such behavior involves the conscious and purposeful biasing of ratings in order to achieve some evaluation-related objective, such as enhancing one's own relative standing. Previous research suggests that these system characteristics may influence the extent to which raters may have an interest in intentionally distorting their ratings of others ([Antonioni, 1994](#); [Toegel & Conger, 2003](#)). In all three cases I focus on team-level, shared perceptions of these attributes. From a social constructionist perspective, it is likely that shared member perceptions of these system characteristics, while certainly influenced by the true nature of these attributes, are likely to have a more direct impact on the rating behaviors of team members than the system characteristics themselves. Moreover, it has been well established that shared perceptions of instituted procedures and practices play an important role in predicting the behavior of groups and their members ([Naumann & Bennett, 2000](#); [Zohar & Luria, 2005](#)).

3.2. Rating purpose and intentional distortion

Since [Meyer, Kay and French \(1965\)](#) first recognized the “dual nature” of performance appraisal, the performance evaluation literature has tended to distinguish between appraisal for developmental and for evaluative or administrative purposes. Studies examining the impact of rating purpose on *supervisory* ratings of subordinates indicate that ratings obtained for administrative purposes (e.g., compensation, promotion decisions) are generally more lenient than those generated for research, feedback or developmental purposes ([Jawahar & Williams, 1997](#)). However, consistent with [Jawahar and Williams' \(1997\)](#) findings regarding subordinate ratings of superiors, [Greguras et al. \(2003\)](#), in a study examining the impact of rating purpose in specifically *peer* evaluation, find rating purpose to have no significant effect on leniency. [Greguras et al. \(2003\)](#) attribute this to the fact that peers are more likely than supervisors to feel that they are in direct competition with their ratees, and so hesitate to evaluate peers leniently out of concern for the potential ramifications if they themselves are given negative evaluations. A second possible explanation is that leniency towards some (e.g., allies) is offset by rating harshness towards others (e.g., adversaries). Drawing from [Greguras et al.](#), I posit that, regardless of management's actual intent in collecting peer evaluation data, the stronger the perception among team members that peer evaluation is being used for administrative purposes, the greater the team-level prevalence of actual rating harshness (i.e., downward intentional distortion).

This issue of team members' perceptions of rating purpose is critical. Regardless of management's espoused intentions (e.g., that evaluation data will be used for developmental purposes only), employees may perceive that the results of a peer assessment are likely to influence administrative decisions having to do with compensation and/or

staffing in the future (Atwater, Waldman, Atwater, & Cartier, 2000) — outcomes that are, more often than not, distributed on a zero-sum basis in organizations. As Tornow and London (1998) suggest, because many organizations fail to adequately communicate the developmental purpose of peer assessment, rater ambiguity over the eventual use of peer assessment data is a relatively common problem. Moreover, as Fedor, Bettenhausen and Davis (1999: 96) note, “even when an organization clearly states how it intends to use performance appraisal information, the inevitable ambiguities associated with administering the system and employees’ efforts to make sense of the entire process can lead employees to perceive differently the extent to which the appraisal system is used for developmental feedback purposes.” These ambiguities may have a basis in fact, as organizations have been known to implement peer evaluation for developmental purposes, only later deciding to use these “rich evaluative data” for administrative decision-making as well. As a result, employees often feel threatened even by developmental evaluation systems (DeNisi & Kluger, 2000: 136).

Consequently, beyond any positive shift in the incentive framework governing *ratee* behavior and job performance, from a social dilemma perspective it is likely that any shift in shared perceptions regarding the purpose of the evaluation will also change the perceived incentive structure governing members’ *own* rating behaviors when evaluating the performance of their peers. Specifically, when peer evaluation data are perceived by team members as being used or likely to be used as a basis for administrative decisions – and in particular, for decisions regarding highly desired, competitive (i.e., zero-sum distributed) organizational rewards such as a promotions or bonuses – it is reasonable to assume that some team members may begin to consider the strategic implications of one evaluation strategy over another (Kozlowski et al., 1998). That is, as soon as team members begin to perceive certain evaluation outcomes as having certain benefits or costs to themselves, they may begin to view the evaluation system as having a more salient incentive structure. Depending on the nature of this perceived incentive structure, team members may be encouraged to intentionally distort their evaluations of their peers, with ratings of those viewed as serious competitors downwardly distorted. This suggests the following proposition:

Proposition 1a. *The stronger team members' shared perceptions – going into a round of peer evaluation – that evaluation data may ultimately be used as the basis for decisions regarding the distribution of valuable and competitive rewards (e.g., compensation), the more pervasive will be intentional downward rating distortion in that round.*

3.3. Rater accountability and intentional distortion

As suggested by Beckner, Highhouse and Hazer (1998: 210), rater accountability concerns the degree to which raters believe themselves to be identifiable, and their ratings to be subject to the review of others. Given the central role of rater identifiability in the accountability construct, anonymous peer evaluation systems, by definition, offer the most limited degree of rater accountability. In contrast, even rating systems in which raters are not explicitly told that their ratings will be subject to review, but in which ratings are nonetheless collected non-anonymously, may be considered to offer a higher degree of accountability, since raters cannot rule out the possibility that their ratings will be reviewed. Similarly, accountability may be enhanced if raters, while promised anonymity before their *ratees*, are still required to justify and sign-off on their ratings, which may be subject to supervisory review.

Although most peer evaluation programs have been adopted on the basis of complete rater anonymity (London, Smither, & Adsit, 1997), the effects of rater accountability on the rating behaviors of *peer* evaluators remain largely unknown. Studies addressing the implications of accountability in multi-source feedback point to an inherent dilemma with respect to the impact of accountability on rater behavior, prompting London et al. (1997:165) to refer to accountability as the “Achilles heel” of multi-source feedback. The dilemma stems from the tendency of *ratees* to prefer that raters be made accountable and to view feedback from accountable raters as less subject to intentional bias, and the obverse tendency of *raters* to prefer anonymity, and to offer more lenient evaluations when they perceive that others (including *ratees*) might be able to identify them as the source of a given rating (Antonioni, 1994).

The empirical literature appears to reflect this dilemma. To date, much of the literature on rater accountability in performance evaluation in general suggests that range restriction and upward intentional distortion (i.e., leniency) tend to be less prevalent in appraisal systems structured around rater anonymity (Klimoski & Inks, 1990). Raters may be hesitant to provide honest and frank negative feedback if they feel that doing so will expose them to potentially high pecuniary costs in the form of *ratee* retaliation in subsequent evaluation rounds, or some sort of administrative sanction. That is, individuals may be willing to accurately report negative impressions only if they feel they will not be asked to

pay a personal price for doing so (Baron & David, 1999). By limiting the ability of administrators and ratees to hold raters accountable for what they view as biased evaluations, anonymous evaluation systems may greatly diminish these perceived pecuniary costs, and thus generate fewer range restriction and leniency problems.

However, most of the studies examining rater accountability have done so in the context of traditional supervisory assessment or upward feedback (e.g., Antonioni, 1994). Findings from such studies may not necessarily be generalizable to peer evaluation in that, as noted earlier, peer raters – unlike supervisors rating subordinates or subordinates rating supervisors – often view themselves in direct competition (for career advancement, compensation, etc.) with those they are rating (Greguras et al., 2003). As a result, what may appear to be an anonymity-induced reduction in rating leniency may in fact be an anonymity-induced tendency towards rating harshness, with non-anonymous ratings perhaps providing a more accurate indication of true performance (Bamberger et al., 2005). In this context, as Mero and Motowidlo (1995) suggest, increased rater accountability may have an overall positive effect on rater accuracy, in that it essentially enhances the degree of alignment between rater and organizational interests in evaluation accuracy. For example, consistent with organizational interests, as accountability increases, raters can be expected to give more thought to rating decisions (Ford & Weldon, 1981). Indeed, Beckner et al. (1998) found that accountable peer raters took a significantly longer time to return their assessments than peer raters evaluating under conditions of anonymity. Furthermore, from a social dilemma perspective, Kozlowski et al. (1998) suggest that enhanced rater surveillance and monitoring are likely to improve rater accuracy in that they diminish the potential benefits to the rater of intentional distortion. For instance, all else being equal, it may be in the rater's self-interest to intentionally distort his or her evaluations of others so as to achieve particular rating objectives, or to compensate for distortions in rating outcomes that are perceived to be inherent to the organizational evaluation system. However, under conditions of increased accountability, the potential costs of distortion may begin to outweigh such benefits.

Indeed, reviews of the social dilemma literature (e.g., Dawes, 1980: 187) suggest that the lack of perceived rater accountability in peer assessment may create an incentive structure in which peer raters increasingly come to believe that it is in their best interest to intentionally distort their assessments of others. While as noted above, rater anonymity may reduce the traditional pecuniary costs to raters of rating accuracy (thereby attenuating any bias toward leniency); it also lowers any disincentive for harshly distorting assessments of others, since it is difficult for peers or appraisal administrators to identify precisely who is intentionally distorting their ratings. And without the ability to identify the intentionally distorting rater, it is impossible for ratee-victims to hold a particular rater accountable, let alone confront or retaliate against him or her. From the rater's perspective, under conditions of anonymity, downwardly distorting a rating is a relatively low-risk activity since there is little risk that the rater will be held accountable or sanctioned; in the worst case, a rating deemed to be an obvious outlier will simply be excluded from the ratee's evaluations. Consequently, *particularly* under conditions of rater anonymity, a shared perception is likely to develop among team members that raters will not be held accountable for their ratings, and thus have little to lose by intentionally and downwardly distorting their evaluations of others.

Moreover, a shared perception of limited rater accountability may in fact increase the incentive for peer evaluators to provide intentionally harsh ratings so as to protect themselves from the intentional distortions expected on the part of their peers. Under conditions of limited perceived rater accountability, raters' fears of strategic downward distortion by their peers, combined with the absence of any reasonable sanctioning mechanism, may lead them to adopt *proactive* downward biasing. That is, consistent with the notion of "backwards induction" (Luce & Raiffa, 1957), raters may adopt the defensive and essentially "adaptive" strategy of proactively and downwardly distorting their assessments of others (Kozlowski et al., 1998). In doing so, they may reduce the risk that others' intentional rating distortion poses to their own overall relative standing. Consequently, while perhaps opening the door to a limited degree of range restriction and leniency bias¹, increasing team members' shared perceptions of rater accountability (e.g., by requiring raters to sign off on their assessments and/or increasing system surveillance) may result in a net decline in actual downward intentional rating distortion. Consequently, I propose:

Proposition 1b. *The stronger team members' shared perceptions of rater accountability going into a round of peer evaluation, the less pervasive will be actual intentional downward rating distortion in that round.*

¹ Limited because it is generally counter to a rater's self-interests to inflate his/her ratings of peers potentially competing for some organizational resource distributed on the basis of a zero sum game.

3.4. Regularity of evaluation and intentional distortion

A third contextual factor potentially influencing the degree to which peer raters engage in intentional rating distortion concerns shared team member perceptions regarding the regularity of peer evaluation. In many organizations, peer evaluation is implemented so infrequently that employees have no sense as to when or even if a future round of evaluation will be conducted. Indeed, London and Smither (1995: 826) report that only 15% of the organizations they surveyed administered peer and other forms of multi-source evaluation on a regular basis. When peer evaluation is conducted irregularly, it is not unreasonable to assume that employees will tend to view each successive round of evaluation as independent of any future round. Given the high degree of temporal uncertainty surrounding peer evaluation in most organizations, even those employees who believe that there are likely to be future rounds may doubt that they will be evaluated by the same set of peers in those future rounds.

Toegel and Conger (2003) note that one likely beneficial effect of such rater independence is a reduced risk of a leniency bias in peer-ratings. However, as in the case of rating purpose, lower ratings can no more be attributed to reduced leniency than to increased rater harshness. That is, from a social dilemmas perspective, rather than reducing the risk of rater leniency, rater independence may actually increase the risk of rater “defection,” or the tendency of raters to emphasize individual rationality and intentionally and downwardly distort their ratings of their teammates’ performance. As Axelrod (1984), Axelrod and Dion (1988) have demonstrated, individuals engaged in a social dilemma involving multiple encounters (e.g., two-person iterated prisoners’ dilemma game) understand that their decisions in one encounter or round might affect the decisions of their peers in subsequent rounds, and therefore tend to adopt increasingly cooperative tactics (i.e., refusing to “snitch”). In contrast, individuals playing in the context of temporal uncertainty (where it is unclear whether there will be additional rounds) are much more likely to defect and take that tactic offering the greatest individual (as opposed to collective) utility. Under these circumstances, in other words, raters may cognitively reduce the probability that any intentional downward distortion on their part will lead to peer retaliation in future rounds. While, as I discuss below, raters can still not completely dismiss the possibility that ratees may retaliate for perceived distortion in the context of their regular work interactions with peer raters, such work-based retaliation is likely to be viewed as entailing considerably greater risk on the part of the retaliating ratee (particularly if the work involves a high degree of outcome interdependence with the rater), and hence deemed to be less likely to actually occur.

From a social dilemma perspective, the perceived inability of ratees to reciprocate or retaliate in subsequent rounds is likely to have two critical consequences. First, the effect on raters’ incentive structures means that those who deem it in their self-interest to intentionally distort their ratings will foresee diminished personal risks in doing so. From a simple cost-benefit analysis, the potential rater costs of downward distortion are significantly lower if there is little risk of subsequent retaliation. As a result, employees asked to evaluate their peers in the context of what is commonly perceived to be a one-time or irregular evaluation program may be more likely to engage in intentional harshness. The second effect, based on the first, takes into account the notion of “backwards induction” noted above. Simply put, while morals and social norms may make it hard for some employees to even consider engaging in intentional rating distortion, these same employees may be willing to do so as a matter of self-defense. That is, if they feel that their peers are likely to take advantage of ambiguity regarding the probability of future peer evaluations and consequently distort their own ratings, these employees may prefer to take pre-emptive action and do to others what they fear others might do to them. In this context, their behavior may be seen as more adaptive than maladaptive (Kozlowski et al., 1998).

Whatever the motivation, the net effect of shared perceptions of evaluation irregularity is likely to be the same: namely, an enhanced likelihood of intentional distortion. This suggests the following proposition:

Proposition 1c. *The stronger team members’ shared perceptions – going into a round of peer evaluation – that peer evaluation is/will be conducted on a regular basis, the less pervasive will be actual intentional downward rating distortion in that round.*

3.5. Intentional distortion and ambient peer-rating distortion

The actual pervasiveness of intentional rating distortion in a team is unlikely to be readily known to team members, since it is obviously not in the interest of raters (particularly those distorting their ratings) to readily disclose such information. However, based on Hackman’s theory of group influences on individuals in organizations (1992), I posit

that team members are likely to develop beliefs regarding both the pervasiveness of such distortion and its implications for their own future rating behaviors on the basis of a combination of their own peer evaluation experiences and those of their teammates. Hackman's (1992: 202) theory suggests that such direct and indirect experiences with rating distortion serve as "ambient stimuli," shaping both evaluation-related beliefs and behaviors of team members. In the same way that direct and indirect experiences with sexual harassment and political influence have been shown to both influence and be influenced by team behaviors (Glomb et al., 1997; Raver & Gelfand, 2005), I posit that *ambient peer-rating distortion* – a team-level property reflecting a summation of the rating-distortion experiences reported by team members – is likely to be influenced by the actual pervasiveness of intentional rating distortion in a team, as well as to influence the pervasiveness of such distortion in subsequent rounds.

3.6. Ambient peer-rating distortion (APRD)

According to Hackman's (1992) theory, ambient stimuli such as APRD are likely to influence team members and team-level outcomes regardless of whether or not all team members personally experience the stimulus. What is of concern is not whether team members share the same perceptions about rating distortion in their own experiences, but rather the extent to which such individual experiences – whether direct or indirect – pervade the group setting and are potentially available to all team members (Raver & Gelfand, 2005). Thus, while shared perceptions of the three peer evaluation system characteristics noted above emerge as a team-level property on the basis of a direct composition model (Chan, 1998: 237) requiring within-group agreement, APRD emerges as a team-level property through an additive composition model (Chan, 1998: 236). Team-level constructs derived on the basis of the additive composition model reflect the summation of the lower level units (i.e., individuals' peer-rating distortion experiences), "regardless of the variance among these units" (Chan, 1998: 236). Consequently, regardless of whether these peer-rating distortion experiences at the team-level are shared or not, a team's level of APRD may be deemed to be higher to the extent that a greater proportion of its members view themselves as having either directly or indirectly (i.e., vicariously) experienced such rating distortion.

As a team-level property, APRD is based on the assumption that team members not only are cognizant of possible intentional distortion in others' ratings of themselves, but are also typically aware of similar experiences with rating distortion on the part of their teammates. Due to the egocentric tendencies of many ratees, intentional distortion on the part of peer raters is unlikely to go unnoticed, particularly when this distortion reflects actual downward bias. Extensive research on self-rating suggests that "individuals have a significantly different view of their own job performance than that held by other people" (Thorton, 1980: 268). Indeed, a meta analysis by Harris and Schaubroeck (1988) found an average corrected correlation of 0.36 between peer and self-ratings (as opposed to a correlation of 0.62 between peer and supervisor ratings). In most cases, the lack of agreement stems from a tendency of individuals to overestimate their own contributions relative to others (Ross & Sicoly, 1979). Saavedra and Kwun (1993: 459) note that such self-enhancement "may be an unavoidable aspect of the peer-rating process in a work group." Unfortunately, the tendency of individuals to rate themselves better than others results in many cases where, as Atwater, Waldman and Brett (2002: 196) note, "feedback recipients receive feedback that is more negative than expected."

Harris and Schaubroeck's meta-analysis found support for two likely explanations for such an egocentric bias, one based on attribution theory, and the other based on the notion of rater defensiveness. In the context of attribution theory, Harris and Schaubroeck's findings suggest that while ratees may attribute good performance to their own behaviors and poor performance to contextual factors, peer raters may be more likely to attribute good performance to contextual conditions and poor performance to the ratee's actual disposition. Even when correcting for range restriction, Harris and Schaubroeck still found much higher agreement between ratings by others (i.e., supervisors and peers) than between self-other ratings (i.e., self-peer, self-supervisor).

This finding suggests that peer ratees may be effectively hypersensitive to downward bias on the part of their peers regardless of the degree to which these peers actually engage in downward intentional distortion. When evaluating their peer evaluation feedback, ratees are likely to view any negative deviation from their self-evaluation – whether accurate or not – as contextually based (Saavedra & Kwun, 1993). While ratees in any appraisal context may attribute the deviation to an evaluator's unintentional failure to adequately take situational constraints into account, they may also – particularly in the case of peer evaluation – lay the blame on intentional attempts by one or more peer evaluators to enhance their own relative evaluations. Such attributions of intentional distortion are likely to be more common in teams in which, on average, the deviations between self- and peer-ratings are larger, and such deviations are likely to be

greater on average for members of teams with a higher prevalence of actual intentional distortion. Furthermore, as the prevalence of actual intentional distortion increases in a team, we would expect a greater number of team members to *directly* experience such deviations and attribute these deviations to intentional distortion on the part of their peer raters.

3.7. Indirect sources of APRD

We might also expect actual prevalence to have significant implications with respect to the vicarious or *indirect* rating distortion experiences of team members. Such a spillover effect is likely in that perceived negative deviations from self-evaluations are likely to confront ratees with an experience that calls into question their self-conceptualizations, implying that, as Pittman and D'Agostino (1985: 120) put it, their “understanding of the world is inadequate.” Because such “subjective uncertainty” (McDonald & Westphal, 2003: 3) tends to evoke distress and negative affect, ratees – in an effort to reestablish a sense of certainty and control over personally important outcomes – are likely to turn to those teammates with whom they feel close in the hope of receiving support, encouragement, or advice. In the process, these teammates are likely to be indirectly exposed to the peer-rating distortion experiences of their colleagues.

While, in theory, coworkers to whom ratees present themselves as victims of intentional distortion could challenge or question these assumptions, the social psychological literature suggests that people generally turn for support to others whose beliefs and perspectives are consistent with their own (Swann, 1996). To the extent that, particularly in professional contexts, people with strong social ties such as those inherent in friendships are normatively expected to affirm rather than question each other's capabilities and professional competencies (Ibarra, 1995), like McDonald and Westphal (2003: 8), we would expect such teammates to “affirm each other's self-serving attributions for performance problems.” Based on this logic, the higher the prevalence of actual rating distortion or the greater the proportion of team members directly experiencing rating distortion, the greater the proportion of team members also likely to be *indirectly* exposed to such incidents and the greater the sense that such rating distortion experiences pervade the team environment.

Taken together, such direct and indirect processes suggest that:

Proposition 2. *The greater the pervasiveness of actual intentional downward rating distortion at the team-level in a given round, the higher the subsequent level of ambient peer-rating distortion.*

3.8. The self-reinforcing effects of APRD over multiple rounds of peer evaluation

As noted above, Hackman's (1992) theory suggests that ambient stimuli in a team environment can have significant individual and team-level consequences. For example, a number of studies suggest that perceived unfairness in evaluation procedures can have significant negative performance, withdrawal and retaliation-related implications at both the individual and team-levels (Shaw, Wild, & Colquitt, 2003). Hackman (1992: 209) suggests that “ambient stimuli realize their effects by influencing members' private and implicit assumptions about what behaviors are appropriate or desirable.” Specifically, by serving as “cues that signal what behaviors are likely to lead to what outcomes” (p. 206), they motivate team members to adopt behavior aimed at achieving or avoiding particular classes of outcomes associated with these stimuli.

In this context, egocentric attributions of intentional peer-rating distortion, as well as the tendency of teammate-supporters to affirm such self-serving attributions, may serve as an initial “primer” of what may, over multiple rounds of peer evaluation, become a self-reinforcing tendency toward intentional distortion in peer-ratings. That is, while actual intentional rating distortion may be limited in initial rounds of peer assessment, the level of ambient peer-rating distortion may still be sufficiently high to encourage team members to adopt more cautious or defensive rating behaviors in subsequent rounds in order to protect themselves from a competitive rating orientation on the part of their peers (Erev & Roth, 1998; O'Leary-Kelly & Newman, 2003). Ambient peer-rating distortion may thus provide a justification for actual downward intentional distortion in subsequent ratings of peers by the ratee. Research in upward feedback suggests that individuals receiving lower-than-expected ratings from others tend to subsequently rate these same others lower on a variety of measures (Atwater, Waldman, & Brett, 2000). To a certain degree, intentional distortion in the rating of one's peers may be viewed as a legitimate, defensive response to a negative peer-rating climate in general, and perceived intentional downward bias by one's peers in particular. In light of such intentional

rating inaccuracy, intentional distortion may serve as the only way to safeguard one's own position relative to others. As noted by Kozlowski et al. (1998: 176): "if rating distortions are the norm, a failure to engage in appraisal politics may be maladaptive." Thus:

Proposition 3. *The higher the level of team ambient peer-rating distortion after a given round of peer evaluation, the greater the team-level prevalence of actual intentional downward rating distortion in subsequent rounds of peer evaluation.*

4. The effects of team ambient peer-rating distortion on cooperation-related team processes

Having described the potentially self-reinforcing nature of intentional distortion in multiple-round peer evaluation, I now turn to the implications of this dynamic with respect to cooperation-related team processes. As depicted towards the right of Fig. 1, I posit that team APRD in any given round may both directly influence cooperation-related processes (resulting in a suppression effect; see path 4a) and moderate the influence of peer evaluation on these processes (resulting in an attenuation effect; see path 4b).

A direct suppression effect is posited for a number of reasons. First, from a social dilemma perspective, it is likely that members of teams with a higher level of APRD will avoid going out of their way to engage in cooperative task behaviors. While peer evaluation is intended to promote such cooperative and team-oriented behaviors, members may find little benefit in continuing to behave accordingly in teams marked by a pervasive sense that ratings fail to accurately reflect such pro-social behaviors. No less, members of such teams may have an incentive to refrain from such behaviors in order to avoid the sense that they are being taken advantage of by their peers, or in other words, becoming a "sucker" (Kerr, 1983). Similar predictions may also be made on the basis of control theory (Hyland, 1988). In addition, members of teams with a higher level of APRD may be more vulnerable to the adoption of tit-for-tat behavior with respect to their teammates (Axelrod, 1984). That is, in addition to retaliating by intentionally distorting their ratings of others in subsequent rounds of peer evaluation (DeNisi et al., 1983), members of high APRD teams may be more inclined to use social loafing as a way to retaliate against those viewed as having been unfair to them in the rating process (O'Leary-Kelly & Newman, 2003). Indeed, Naumann and Bennett (2000) found that group procedural justice climate can explain the variance in group member helping behavior.

Second, a high level of team APRD may have a direct negative impact on cooperation-related team processes in that members of such teams may be less inclined to trust the peer evaluation data they receive. Having less faith in the validity of such ratings overall, team members may be more inclined to broadly dismiss all peer feedback data, including unbiased and accurate feedback critical to effective team performance.

Finally, a high level of team APRD may induce affective or relationship conflicts, as hypotheses regarding the source of such biased ratings are developed and implicit mutual incriminations emerge (DeNisi et al., 1983). Perceptions of intentional peer-rating distortion may also breed suspicion, with team members viewing all pro-social or cooperative behavior by their peers as manipulative and intended to "trick" them into becoming a "sucker" (i.e., by providing an alter with a strong evaluation while allowing themselves to be subject to strategic rating by that same alter). The view that such cooperation by others is simply an attempt to "win points" may breed a sense of resentment and, ultimately, affective conflict, particularly if – for example – members begin to suspect one another of passing along high-visibility but relatively useless knowledge (for "image" purposes) while keeping less visible but more valuable information to themselves. A number of team-level studies have found that such types of affective or relationship conflicts can be detrimental to helping and knowledge-sharing team processes, in that team members engaged in such conflict tend to simply avoid one another (Amason, 1996). Similarly, Axelrod (1984) suggests that when peers begin to suspect that cooperative behaviors are opportunistically driven (i.e., motivated by point-grabbing), pre-existing meta-norms of cooperation may be significantly weakened. Consequently, I propose:

Proposition 4a. *The higher the team APRD, the lower team-level task- and knowledge-sharing, and peer-to-peer helping.*

In addition to such direct effects, in teams characterized by a higher level of APRD any direct, positive impact of peer evaluation on cooperation-related team processes team may be attenuated. While the subjective expected utility of behaviors associated with task-sharing, knowledge sharing, and help giving is likely to increase with the enhanced monitoring of such behaviors provided by peer evaluation, over multiple rounds this perceived utility function is likely

to be attenuated in teams marked by a pervasive sense that peer-ratings are subject to intentional downward distortion. Under such conditions, team members may feel that, regardless of the degree to which their behavior is open to monitoring and sanctioning, they have little to gain by cooperating, and little to lose by “defecting” and adopting less cooperative patterns of behavior. If cooperative, team-oriented behaviors are perceived to go unrecognized by biased peer raters, then peer evaluation is unlikely to motivate such behaviors, and peer-rating is less likely to yield the process-enhancing benefits suggested by the peer evaluation literature discussed earlier. Consequently:

Proposition 4b. *Team APRD will moderate the association between peer assessment and cooperation-related team processes such that the higher the team APRD level, the more attenuated the generally positive association between peer evaluation and cooperation-related team processes.*

5. Discussion

The model presented above is unique in the research literature on peer evaluation in that, rather than focusing on individual-level psychometric properties or user acceptability, its team-level orientation allows for a focus on the implications of peer evaluation on team processes, and the social dilemma-like dynamics underlying the emergence of these effects over time. Adopting such an approach, one has little choice but to examine peer-rating as a conscious behavior little different from those behaviors being evaluated, and as such, a phenomenon that is subject to learning processes and implicit incentive structures (Erev & Roth, 1998).

5.1. Theoretical contributions

Associated with the resulting framework are two major contributions to the peer evaluation literature. First, unlike much of the existing literature on peer assessment and multi-source feedback, where the focus is on identifying sources of unintentional distortion and estimating the magnitude of their effects, I highlight the important but under-researched area of intentional rating distortion. Despite recent recognition of intentional rating distortion as a significant issue in supervisory-based performance assessment (Kozlowski et al., 1998), research examining intentional distortion in peer evaluation is rare. As suggested in the discussion above, the relative neglect of intentional distortion in peer assessment is surprising, given the potentially high degree of competition among peer raters as well as the interdependence of peer-ratings over time. Of course, organizational politics and the pecuniary costs of evaluation may also motivate supervisors to intentionally distort subordinate ratings (Kozlowski et al., 1998; Longenecker, Gioia, & Sims 1987). However, given the fact that a rating may be considered meaningful only in relation to the ratings of others, and given that peer raters may also be competitors for valued organizational resources, the motivation for raters to intentionally distort their evaluations of others is likely to be at least as great, if not greater, than a supervisor’s motivation to distort his evaluation of a subordinate. The model laid out above thus contributes to the literature by providing some insights into the possible roots of intentional distortion in peer evaluation, as well as into how initially-weak tendencies towards intentional distortion may become self-reinforcing and increase in strength over time.

A further contribution is made by using a social dilemma framework as a lens through which to analyze the implicit dilemmas that peer raters may be increasingly likely to face over multiple rounds of evaluation. The literature on social dilemmas neatly captures the complex set of pressures faced by peer raters who, on the one hand, are likely to have a common interest in providing honest and helpful feedback to one another, while on the other may be forced to take into account the competitive implications of providing peers with one rating profile over another. By focusing on the implicit learning processes and incentive structures influencing the rating behaviors of peer evaluators, the social dilemma approach provides a dynamic framework for explaining the inconsistent findings in peer evaluation research to date. From a social dilemma perspective, the contradictory findings of different studies on the implications of peer evaluation for team processes and performance may be largely explainable if we consider team members’ shared understandings about the peer evaluation system being implemented, and the impact of these shared understandings on the ambient team context framing peer-rating activity. A social dilemma perspective suggests that any positive consequences after one round of peer evaluation may be reduced or even reversed over subsequent rounds, as team members respond to the ambient team environment in which the evaluation is conducted. Studies finding largely positive consequences for peer evaluation seem to be largely single-round studies (e.g., Erez et al., 2002), and so do not take into account that intentional distortion may be behavior that is learned and reinforced over multiple rounds of peer

evaluation. Conversely, studies finding largely negative consequences appear to either consider peer evaluation over multiple rounds (e.g., Bamberger et al., 2005) or to examine peer evaluation systems with inherently competitive incentive structures (e.g., systems in which results are commonly assumed by team members to be used as a basis for reward-related decisions; e.g., DeNisi et al., 1983), generating an ambient rating environment pervaded by a sense of downward distortion already in the initial round.

5.2. Testing the model

Despite these contributions, given the inherent difficulties in capturing intentional rating distortion, as well as the complex and process-dependent nature of the model, researchers are likely to face a number of challenges in testing the propositions explicated above. The first challenge is to operationalize actual intentional distortion with an eye toward capturing the variance in the pervasiveness of such distortion across teams. While social dilemma researchers have, in an experimental context, been able to rather precisely operationalize distortion-type behaviors, this may be more difficult to accomplish in a field context, given the relative lack of external validity criteria. To date, few attempts at developing or applying such a measure have been reported. One exception is Kozlowski et al. (1998), who analyzed comparative evaluations of navy personnel, focusing in particular on the tendency of supervisor raters to inflate the reported size of a subordinate's comparison group (relative to the actual cohort size) with phantom peers (so as to be able to raise the resulting ranking), and/or to under-report the number of peers below the rate. More recently, Greguras et al. (2003) employed generalizability theory to estimate sources of systematic variability associated with peer-ratings.

An alternative approach – requiring the collection of both self- and peer-assessment data – might be to estimate the likelihood of actual intentional distortion based on an algorithm comparing the relative shifts in a rater's self- and peer-ratings to the mean shift in peer-based ratings of the same targets over two or more evaluation rounds (Bamberger & Bar Niv, 2006). Higher likelihood scores would be generated if, for example, between the first and second (post-feedback) rounds of peer evaluation self-ratings of “ego” were to increase dramatically (both on an absolute level and relative to the mean change in self-ratings of the team), while ego's rating of an “alter” were to decline relative to the mean upward shift in that alter's peer-ratings over the two rounds. The algorithm could also conceivably be expanded to take into account rating patterns indicative of coalition-based intentional distortion targeted against one or more team members (e.g., parallel shifts in *both* self- and alter-specific peer-ratings across two or more coalition partners).

The second challenge involves the testing of a process-dependent, cybernetic model incorporating latent variables grounded on what Chan (1998) refers to as direct consensus constructs. Recent advances in multi-level modeling (Raudenbush & Bryk, 2002) and growth curve modeling (Chan, 2003) greatly facilitate the testing of process-dependent models grounded on repeated measures from a given set of observations (i.e., teams) over time. Such models allow the researcher to precisely measure how relationships among a particular set of variables change over time. The increasing prevalence of multi-level analysis in organizational research has also created a standard (based on mean or median r_{WG} as well as ICC scores) for interpreting aggregated unit-member perceptions as direct consensus constructs or latent, team-level variables (Zohar & Luria, 2005). Nevertheless, the complexity of the model posited above suggests that it may be more prudent to apply these analytical techniques to individual model elements on a step-by-step basis rather than trying to use these methods to test the model as a whole.

As a first step, I recommend that researchers first attempt to establish the degree to which higher levels of team APRD both directly and indirectly affect cooperation-related team outcomes (Propositions 4a and 4b). To ensure a sufficient degree of variance in APRD, researchers might want to collect data from teams in multiple organizations, perhaps using multi-level analyses to take into account any random effects of teams' organizational affiliation.

To the degree that Propositions 4a and 4b are supported, the next step would involve explaining the variance in APRD (Propositions 1a 1b 1c 2) and, in particular, its change over time (Propositions 2 and 3). In order to ensure variance in APRD across groups, it may again make the most sense to draw data from teams nested in multiple organizations, with data regarding members' shared perceptions of particular peer evaluation system characteristics (e.g., accountability, regularity) and APRD collected at two distinct points in time (e.g., system characteristics prior to evaluation and APRD after participants have received evaluation feedback). Using such data, a multi-level mediation model (again controlling for the random effects of teams' organizational affiliation) might be applied to assess the degree to which any direct relationships between team members' shared perceptions of system characteristics (e.g., accountability, regularity) and APRD are mediated by the pervasiveness of actual intentional distortion at the team-level (Propositions 1a 1b 1c 2). Finally, turning to Proposition 3, latent growth-curve modeling might be used to

examine the cybernetic effects of the pervasiveness of actual intentional distortion at the team-level on team APRD over time, and to examine how the relative effects of shared perceptions of system characteristics versus APRD on the pervasiveness of actual intentional distortion at the team-level change over time.

Nevertheless, even following such a step-by-step approach, researchers are likely to encounter a number of problems in empirically testing the propositions generated above. Some of these problems have to do with the less-than-universal applicability of several assumptions implicit in a social dilemma approach to peer evaluation contexts. Others have to do with the variability of peer evaluation programs across organizations.

Regarding the first set of problems, as noted above, a key characteristic of a social dilemma is the conflict between individual and collective rationality (Dawes, 1980). However, this conflict is likely to generate the predicted effects only if it is perceived as such by those team members subject to peer evaluation, and this is not something that can necessarily be taken for granted. While unbiased performance ratings are theoretically a public good, the benefits of which all team members theoretically enjoy regardless of their contribution (Olson, 1965), in some organizational contexts they may not necessarily be recognized as such by team members. Rather, performance ratings and their applications may be poorly understood by employees, with these “naïve” peer raters, at least initially, seeing little direct association between the ratings that they provide and any direct advantage or disadvantage to themselves.

The peer evaluation context may also violate other core assumptions underlying the public goods paradigm suggested by Olson (1965) if supervisors are able to link particular assessments with particular peer raters. For example, this paradigm suggests that all team members share equal power in determining evaluation outcomes. However, while formally an individual’s summary score is typically the mean of the ratings provided by all of their rating peers, in cases where organizations use peer evaluation data as a basis for administrative decisions, supervisors may informally weight some peers’ ratings more heavily than others’. Other assumptions that may be violated include the notion that all team members have an equal payoff or penalty when teammates cooperate or defect, and that all team members perform their ratings simultaneously and independently. The equal payoff/penalty assumption may be violated if peer-ratings are considered in light of ratings provided by other sources (e.g., supervisor, subordinates) with, for example, more severe penalties for low ratings among those receiving low scores from other sources. The simultaneity/independence assumption may be violated if conditions allow peers to “make deals” or form rating coalitions with one another.

However, even in such cases, peer evaluators are still likely to experience the same basic conflict between cooperative versus competitive rating strategies. That is, even in the absence of these boundary conditions, it is still highly likely that, lacking the *cooperative* participation of all team members in the peer evaluation enterprise, the effort may generate social “side effects” that may counter-balance many of its intended benefits. Moreover, while the public good paradigm presented above may not apply in all peer evaluation contexts, Erev and Rappaport (1990) have suggested other models (e.g., sequential social dilemmas) with more relaxed assumptions that might be applicable.

Regarding the second (variability) issue, while there may be variability across organizations on each of the three peer evaluation system characteristics individually, the number of organizations employing interdependent teams that conduct *multiple rounds of evaluative* peer evaluation on a strictly *anonymous* basis may be limited. Indeed, London and Smither (1995: 805) report that only 40% of all firms that they surveyed conducted administered multi-source feedback on two or more occasions. As such, while some of the strongest team-level effects might be expected when testing for two- or three-way interactions among the system characteristic variables, such testing may still have limited practicality in many organizations.

6. Conclusion

Despite these potential empirical limitations, the model presented above represents the first attempt to examine the potential longer-term, team-level implications of peer evaluation when such systems are adopted in interdependent teams. While some of the unexplained variance in peer evaluation may result from yet-to-be-identified forms of unintentional bias, the propositions generated above provide some initial insight into how shared, team-member perceptions of system characteristics may shape the team-level prevalence of intentional distortion in peer-ratings as well as the team’s ambient rating environment, and how these may influence cooperation-related team processes over time. Clearly, this model presents only a very simple framework for what is likely to be a far more complex dynamic. For example, the effects of the system characteristics discussed above are likely to be moderated by a variety of team characteristics, such as the degree to which the team is characterized by a competitive climate, or the extent of outcome

interdependence (Wageman, 1995), team size, and team cohesion (O'Leary-Kelly, 1998). As such, researchers should be encouraged to both test and expand upon this model. As noted by Kozlowski et al. (1998: 190): "This type of intentional evaluative strategizing and its effects represents an entire class of organizational behavior that remains little explored in the performance appraisal literature."

References

- Albanese, R., & Van Fleet, D. (1985). Rational behavior in groups: The free-riding tendency. *Academy of Management Review*, *10*, 244–255.
- Amason, A. (1996). Distinguishing the effects of functional and dysfunctional conflict on strategic decision making: Resolving a paradox for top management teams. *Academy of Management Journal*, *39*, 123–148.
- Antonioni, D. (1994). The effects of feedback accountability on upward appraisal ratings. *Personnel Psychology*, *47*, 349–356.
- Atwater, L. E., Waldman, D., Atwater, D., & Cartier, P. (2000). An upward feedback field experiment: Supervisors' cynicism, reactions, and commitment to subordinates. *Personnel Psychology*, *53*, 275–297.
- Atwater, L. E., Waldman, D. A., & Brett, J. F. (2000). *The effect of upward feedback on Leader Member Exchange (LMX) and liking relationships*. Working paper Phoenix, AZ: Arizona State University.
- Atwater, L. E., Waldman, D. A., & Brett, J. F. (2002). Understanding and optimizing multisource feedback. *Human Resource Management*, *41*, 193–208.
- Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic.
- Axelrod, R., & Dion, D. (1988). The further evolution of cooperation. *Science*, *242*, 1385–1390.
- Bamberger, P. A., & Bar Niv, O. (2006). Intentional Rating Distortion and Peer Evaluation in Management Education: Why and How to Identify "Game-Players". *Journal of Learning in Higher Education*, *2*, 77–87.
- Bamberger, P., Erev, I., Kimmel, T., & Oref-Chen, T. (2005). Peer assessment, individual performance and contribution to group processes: The impact of rater anonymity. *Group and Organization Management*, *30*, 344–377.
- Baron, J., & David, K. (1999). *Strategic Human Resource Management*. New York: Wiley.
- Baron, J. N., & Kreps, D. M. (1999). *Strategic human resources: Frameworks for general managers*. New York: John Wiley.
- Beckner, D., Highhouse, S., & Hazer, J. T. (1998). Effects of upward accountability and rating purpose on peer-rater inflation and delay: A field experiment. *Journal of Organization Behavior*, *19*, 209–214.
- Bernardin, H. J., & Villanova, P. J. (1986). Performance appraisal. In E. A. Locke (Ed.), *Generalizing from laboratory to field settings* (pp. 43–62). San Francisco: New Lexington Press.
- Bracken, D. W., Timmerck, C. W., & Church, A. H. (2001). *The handbook of multisource feedback*. San Francisco: Jossey-Bass.
- Campion, M. A., Medsker, G. J., & Higgs, A. C. (1993). Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel Psychology*, *46*, 823–850.
- Cederblom, D., & Lounsbury, J. W. (1980). An investigation of user acceptance of peer evaluations. *Personnel Psychology*, *33*, 567–579.
- Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology*, *83*, 234–236.
- Chan, D. (2003). Data analysis and modeling longitudinal processes. *Group and Organization Management*, *28*, 341–365.
- Cohen, S. G., & Bailey, D. E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management*, *23*, 239–290.
- Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology*, *31*, 169–193.
- DeNisi, A. S., & Kluger, A. (2000). Feedback effectiveness: Can 360-degree feedback appraisals be improved? *Academy of Management Executive*, *14*, 129–139.
- DeNisi, A. S., Randolph, W. A., & Blencoe, A. (1983). Potential problems with peer ratings. *Academy of Management Journal*, *26*, 457–464.
- Drexler, J. A., Beehr, T. A., & Stetz, T. A. (2001). Peer appraisals: Differentiation of individual performance on group tasks. *Human Resource Management*, *40*, 333–345.
- Druskat, V. U., & Wolff, S. B. (1999). Effects and timing of developmental peer appraisals in self-managing work groups. *Journal of Applied Psychology*, *84*, 58–74.
- Erev, I., & Roth, A. (1998). Predicting how people play games: Reinforcement learning processes on games with unique mixed strategy equilibrium. *American Economic Review*, *88*, 848–881.
- Erez, A., Lepine, J. A., & Elms, H. (2002). Effects of rotated leadership and peer evaluation on the functioning and effectiveness of self-managed teams: A quasi-experiment. *Personnel Psychology*, *55*, 929–948.
- Fedor, D. B., & Bettenhausen, K. L. (1989). The impact of purpose, participant preconceptions and rating level on the acceptance of peer evaluations. *Group and Organization Studies*, *14*, 182–197.
- Fedor, D. B., Bettenhausen, K. L., & Davis, W. (1999). Peer reviews: Employees' dual role as raters and recipients. *Group and Organization Management*, *24*, 92–120.
- Fisher, J. D., Goff, B. A., Nadler, A., & Chinsky, J. M. (1988). Social psychological influences on help seeking and support from peers. In B. H. Gottlieb (Ed.), *Marshaling social support: Formats, processes and effects* (pp. 267–304). Newbury Park, CA: Sage.
- Ford, J. K., & Weldon, E. (1981). Forewarning and accountability: Effects on memory-based interpersonal judgments. *Personality and Social Psychology Bulletin*, *7*, 264–268.
- Glomb, T. M., et al. (1997). Ambient sexual harassment: An integrated model of antecedents and consequences. *Organizational Behavior and Human Decision Processes*, *71*, 309–328.

- Greguras, G. J., Robie, C., Schleicher, D. J., & Goff III, M. (2003). A field study of the effects of rating purpose on the quality of multisource ratings. *Personnel Psychology*, *56*, 1–21.
- Hackman, J. R. (1987). The design of work teams. In J. W. Lorsch (Ed.), *Handbook of organizational behavior* (pp. 315–342). Englewood Cliffs, NJ: Prentice-Hall.
- Hackman, J. R. (1992). Group influences on individuals in organizations. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial organizational psychology*. Palo Alto, CA: Consulting Psychologists Press.
- Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self–boss, self–peer, and peer–boss ratings. *Personnel Psychology*, *41*, 43–62.
- Hazucha, J. F., Hezlett, S. A., & Schneider, R. J. (1993). The impact of 360-degree feedback on management skills development. *Human Resource Management*, *32*, 325–351.
- Higgins, M. C., & Kram, K. E. (2001). Reconceptualizing mentoring at work: A developmental network perspective. *Academy of Management Review*, *26*, 264–288.
- Hyland, M. E. (1988). Motivational control theory: An integrative framework. *Journal of Personality and Social Psychology*, *55*, 642–651.
- Ibarra, H. (1995). Race, opportunity, and diversity of social circles in managerial networks. *Academy of Management Journal*, *38*, 673–703.
- Jawahar, I. M., & Williams, C. R. (1997). Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology*, *50*, 905–925.
- Kane, J., & Lawler, E. (1978). Methods of peer assessment. *Psychological Bulletin*, *85*, 555–586.
- Kerr, N. L. (1983). Motivation losses in small groups: A social dilemma analysis. *Journal of Personality and Social Psychology*, *45*, 819–828.
- Kerr, N. L., & Brunn, S. (1981). Ringelmann revisited: Alternate explanations for the social loafing effect. *Personality and Social Psychology Bulletin*, *7*, 224–231.
- Klimoski, R., & Inks, L. (1990). Accountability forces in performance appraisal. *Organizational Behavior and Human Decision Processes*, *45*, 194–208.
- Kozlowski, S. W., Chao, G. T., & Morrison, R. F. (1998). Games raters play: Politics strategies and impression management in performance appraisal. In J. W. Smither (Ed.), *Performance appraisal: State of the art in practice* (pp. 163–205). San Francisco: Jossey-Bass.
- Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management*, *30*, 881–905.
- London, M., & Smither, J. (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*, 803–839.
- London, M., Smither, J., & Adsit, D. (1997). Accountability: The Achilles heel of multi-source feedback. *Group and Organization Management*, *22*, 162–184.
- Longenecker, C., & Ludwig, D. (1990). Ethical dilemmas in performance appraisals revisited. *Journal of Business Ethics*, *9*, 961–969.
- Longenecker, C. O., Gioia, D. A., & Sims, H. P. (1987). Behind the Mask: The Politics of Employee Appraisal. *The Academy of Management Executive*, *1*, 183–193.
- Love, K. G. (1981). Comparison of peer assessment methods: Reliability, validity, friendship bias and user reaction. *Journal of Applied Psychology*, *66*, 451–457.
- Luce, R., & Raiffa, H. (1957). *Games and decisions*. New York: Wiley.
- McDonald, M. L., & Westphal, J. D. (2003). Getting by with the advice of their friends: CEOs' advice networks and firms' strategic responses to poor performance. *Administrative Science Quarterly*, *48*, 1–32.
- Mero, N. P., & Motowidlo, S. J. (1995). Effects of rater accountability on the accuracy and the favorability of performance ratings. *Journal of Applied Psychology*, *80*, 517–524.
- Meyer, H. H., Kay, E., & French, J. R. P. (1965). Split roles in performance appraisal. *Harvard Business Review*, *43*, 123–129.
- Miller, J. S., & Cardy, R. L. (2000). Self-monitoring and performance appraisal: Rating outcomes in project teams. *Journal of Organization Behavior*, *21*, 609–626.
- Mitchell, T. R., & Liden, R. C. (1982). The effects of the social context on performance evaluation. *Organizational Behavior and Human Decision Processes*, *40*, 149–169.
- Mohr, L. (1982). *Explaining organizational behavior*. San Francisco: Jossey-Bass.
- Murphy, K. R., & Cleveland, J. (1991). *Performance appraisal*. Needham Heights: Allyn and Bacon.
- Murphy, K. R., & Cleveland, J. (1995). *Understanding performance appraisal: Social, organizational and goal-based perspectives*. Thousand Oaks, CA: Sage.
- Nadler, A. (1998). Relationship, esteem and achievement perspectives on autonomous and dependent help seeking. In S. A. Karabenik (Ed.), *Strategic help seeking: Implications for learning and teaching*. Mahwah: L. Erlbaum.
- Napier, N. K., & Latham, G. P. (1986). Outcome Expectancies of People Who Conduct Performance Appraisals. *Personnel Psychology*, *39*, 827–837.
- Naumann, S. E., & Bennett, N. (2000). A case for procedural justice climate: Development and test of a multilevel model. *Academy of Management Journal*, *43*, 881–889.
- O'Leary-Kelly, A. (1998). The influence of group feedback on individual group member responses. *Research in Personnel and Human Resource Management*, *16*, 255–294.
- O'Leary-Kelly, A., & Newman, J. (2003). The implications of performance feedback research for understanding antisocial work behavior. *Human Resource Management Review*, *13*, 605–629.
- Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups*. Cambridge, MA: Harvard Univ. Press.
- Pittman, T. S., & D'Agostino, R. R. (1985). Motivation and attribution: The effects of control deprivation on subsequent information processing. In J. H. Harvey & G. Weary (Eds.), *Attribution: Basic issues and applications* (pp. 117–141). Orlando: Academic Press.
- Raver, J. L., & Gelfand, M. J. (2005). Beyond the individual victim: Linking sexual harassment, team processes and team performance. *Academy of Management Journal*, *48*, 387–400.

- Ross, M., & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology*, 37, 322–336.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*, 2nd Edition. Thousand Oaks, CA: Sage.
- Saavedra, R., & Kwun, S. (1993). Peer evaluation in self-managing work groups. *Journal of Applied Psychology*, 78, 450–462.
- Schneider, B. (2000). The psychological life of organizations. In N. M. Ashkanasy, D. P. M. Wilderom, & M. F. Peterson (Eds.), *Handbook of organizational culture and climate* (pp. xvii–xxii). Thousand Oaks, CA: Sage.
- Shaw, J. C., Wild, E., & Colquitt, J. A. (2003). To justify or excuse?: A meta-analytic review of the effects of explanations. *Journal of Applied Psychology*, 88, 444–458.
- Swann Jr., W. B. (1996). *Self-traps: The elusive quest for higher self-esteem*. New York: Freeman.
- Thorton, G. (1980). Psychometric properties of self-appraisal of job performance. *Personnel Psychology*, 33, 262–271.
- Toegel, G., & Conger, J. A. (2003). 360-degree assessment: Time for reinvention. *Academy of Management Learning and Education*, 2, 297–311.
- Tornow, W. W., & London, M. (1998). Challenges and implications for maximizing 360-degree feedback. In W. W. Tornow & M. London (Eds.), *Maximizing the value of 360-degree feedback* (pp. 249–258). San Francisco: Jossey-Bass.
- Wageman, R. (1995). Interdependence and group effectiveness. *Administrative Science Quarterly*, 82, 145–280.
- Zohar, D., & Luria, G. (2005). A multilevel model of safety climate: Cross-level relationships between organization and group-level climates. *Journal of Applied Psychology*, 90, 616–628.