12 RATER GOALS AND RATING DISTORTION

Learning Objectives

- 12.1 Understand why rating inflation is more likely to be the result of rater motivation than of rater errors
- 12.2 Learn about the goals raters are likely to pursue when rating their subordinates
- 12.3 Learn how the expected outcomes of giving accurate versus inflated ratings can shape rater behavior
- 12.4 Consider the roles of individual versus contextual factors in explaining rating distortion

In Chapter 1, we noted that performance ratings appear inflated in most organizations; a very large percentage of the workers who receive performance appraisals receive ratings that are well above "average," and it is not unusual to find 80% of all performance appraisal ratings to be at or near the top end of the rating scale. This is not a recent phenomenon; Bjerke, Cleveland, Morrison, and Wilson (1987) report that evaluations of naval officers, going back as far as the 1860s, show evidence of inflation, with most officers receiving ratings of above average or excellent. Similarly, this is not a phenomenon that is unique to North American organizations, but rather it is seen in many countries (e.g., several recent papers describe rating inflation in Asian organizations, including Barron & Sackett, 2008; Ng, Koh, Ang, Kennedy, & Chan, 2011).

The fact that ratings are high does not automatically mean that they are inaccurate. It is always possible that a rater who assigns high ratings to his or her subordinates is simply recognizing that these employees *are* good performers. There is evidence, however, that raters do indeed distort performance ratings, consciously choosing to give their subordinates higher evaluations than they deserve. For example, more than 70% of managers in Longenecker and Ludwig's (1990) study admitted giving their subordinates ratings that did not reflect their actual level of performance. Usually, this involves inflating ratings, but in some cases raters deliberately gave their subordinates unduly low ratings, often to send a message to poor performers that improvements were needed.

There are multiple lines of evidence suggesting that deliberate distortion is more prevalent than unintentional bias (Bernardin & Villanova, 1986). First as noted above, when supervisors and managers are interviewed about the ratings they assign, they readily admit giving their subordinates inflated ratings. There is also indirect evidence of conscious distortion. For example, making raters more accountable has been shown to lead to lower performance ratings (Bamberger, 2007; Mero & Motowidlo, 1995). The usual interpretation of this finding is that when raters are held accountable, they realize that they will not get away with distortion, and therefore give more accurate ratings.

Most of the research on distortion in performance ratings focuses on supervisory ratings of performance, but there are reasons to believe that distortion also occurs when giving upward feedback (e.g., rating the performance of your superiors). This sort of upward feedback is a popular component of 360-degree rating systems, but the process of giving performance feedback to one's superiors is a potentially risky one. Much like supervisory rating of their subordinates, employees who are asked to provide upward feedback may be motivated to give inflated ratings. This motivation depends in large part on the employees' perceptions of the benefits of giving honest feedback versus the risk of retaliation if they give low ratings or critical feedback (Smith & Fortunato, 2008). The motivation to provide honest versus inflated feedback is also likely to be influenced by the employee's degree of trust versus cynicism regarding the motives of top management and the reasons for seeking upward feedback, as well as their belief that they have the information, knowledge, and skills needed to accurately evaluate the performance of their superiors.

Why Are Ratings Inflated?

For many years, the shortcomings of performance appraisal systems were assumed to reflect problems with the rating scales or with rater training, and it was believed that the difficulty in providing accurate and useful performance ratings could be reduced if we gave raters better tools and skills. The flaw in this line of thinking is that it treated raters as passive measurement instruments, and assumed that raters were always trying their best to provide accurate measures of each ratee's performance. Starting in the 1980s, several authors (e.g., Banks & Murphy, 1985; Cleveland & Murphy, 1992; Longenecker, Sims, & Gioia, 1987; Murphy & Cleveland, 1991, 1995; Spence & Keeping, 2011) suggested that some of the apparent failures of performance appraisal could be better explained in terms of motivation than in terms of ability and skill. That is, it is wrong to assume that the reason that managers and supervisors fail to provide accurate and useful performance evaluations is that they *cannot* accurately evaluate performance. Rather, it appears that raters often appear to give ratings that are inflated, or ratings that fail to discriminate among their subordinates, because they prefer giving inflated or distorted ratings to giving accurate performance ratings.

The critical breakthrough in research and theory came as a result of something researchers too often forget to do. Starting in the 1980s, several researchers (e.g., Bjerke et al., 1987; Longenecker et al., 1987) spent time systematically sitting down with supervisors, managers, and others involved in performance appraisal, talking with them about what they were thinking and about what they were trying to accomplish when completing performance appraisal forms, and paying attention to what they said. Two themes emerged from these conversations. First, very few people said, "I am trying my best to provide accurate appraisals, but just cannot seem to manage it." Second, the raters these researchers talked with almost always had sensible rationales for giving ratings that appeared to be inflated or that failed to distinguish between employees, even when the differences in their performance levels seemed fairly obvious. These explanations usually involved using performance appraisal as a tool for accomplishing some important goals (e.g., motivating employees) rather than using performance appraisal as a simple measurement instrument.

It is important to keep in mind that the fact that people *say* that they are pursuing some goals and not pursuing others does not necessarily mean that these goals actually drive their behavior. However, their consistent claim that they had good reasons for giving performance ratings that were not accurate did stimulate the development of models of the appraisal process that gave prominent attention to the goals and intentions of raters (e.g., Cleveland & Murphy, 1992), and these models have received considerable empirical support (Murphy, Cleveland, Skattebo, & Kinney, 2004; Murphy, Deckert, & Hunter, 2013; Wang, Wong, & Kwong, 2010; Wong & Kwong, 2007).

In our opinion, a consistent pattern of rating inflation, which is seen across many organizations and across many cultures, can only be adequately understood in terms of the goals and strategies raters adopt when completing performance appraisal forms. After all, it is unlikely that raters do not *know* that some of their subordinates actually perform better than others. One of our professors, Frank Landy, enjoyed telling a story about an interview with the fire chief of a major city, who claimed that there were no differences in the performance of his firefighters, and that they were all equally excellent. Frank then asked him if he would care who responded if his own house was on fire, and of course he had no difficulty telling Frank which engine and which firefighters he would want to come to try and save *his* house. If managers *know* that some of their subordinates are in fact average or poor performers, why do they consistently give them high performance ratings? Goal-directed models propose that ratings are inflated because raters believe that a number of important ends are more likely to be achieved if ratings are inflated than if they are accurate (Cleveland & Murphy, 1992; Murphy & Cleveland, 1991, 1995).

In this chapter, we discuss the nature and content of the goals that might direct the rater's behavior. Early work on rater goals was largely anecdotal (Bjerke et al., 1987; Longenecker et al., 1987); a number of more recent studies have helped in identifying the goals raters pursue and have confirmed that raters who pursue different goals often end up assigning different performance ratings. For example, Spence and Keeping (2011) reviewed several studies of the goals raters pursue when completing performance ratings, and suggested that the factors that influence a rater's likelihood of providing inflated ratings included: (1) avoiding confrontations with subordinates, (2) the reluctance to take time away from other valued tasks to provide accurate ratings, (3) a desire to be in compliance with organizational norms, (4) the need to promote problem employees out of a department, (5) the desire to look like a successful and competent manager, and (6) to motivate and act in the best interest of the employee. This list is far from exhaustive, but it does give an indication of the range of considerations that may weigh on raters when they complete performance appraisal forms.

Whose Goals?

It is useful to distinguish between the purpose of performance, as articulated by the organization, and the rater's goals when completing a performance appraisal. Cleveland, Murphy, and Williams (1989) note that the organization often imposes, or attempts to impose, a number of goals on the rater that are linked to the way they use performance appraisal. In particular, they suggested that appraisals that are designed for developmental purposes should differ in important ways from appraisals designed to be used to determine pay or promotions. (This idea is far from original; it dates back to at least Meyer, Kay, & French, 1965.) If appraisals are used primarily for developmental purposes, raters should be focusing on the individual strengths and weaknesses of individual employees, and should pay less attention to their overall performance levels. On the other hand, if appraisals are used to make administrative decisions (e.g., merit pay, promotion, layoffs), raters should be focusing on rank-ordering employees in terms of their overall effectiveness, and should pay less attention to within-ratee variability in performance. You might notice that we use the term *should* here. The question of whether raters actually *will* use the appraisal from the way the organization wants them to will depend largely on whether or not they accept and internalize the goals set out by the organization. That is, if the organization proposes using appraisal to identify strengths and weaknesses, the likelihood of obtaining the sort of developmental information that is desired (e.g., identification of strengths and weaknesses) is higher if raters accept this goal as legitimate and important.

Similarly, organizations impose performance goals on individuals, work groups, and larger organizational units (Cheng, Luckett, & Mahama, 2007). While these goals are important for determining how organizations evaluate the effectiveness of these individuals or groups, such externally imposed goals and metrics are distinct from the sort of rater goals this chapter describes. When we use the term *goals*, we refer to the objectives both raters and ratees seek when they participate in the performance appraisal process. Thus, the organization's policies, their statements about the purpose of appraisal, and the goals or targets organizations set are likely to have quite different impacts on the behavior of raters and ratees when they are consistent with the rater's own objectives than when they are not (Erez & Kanfer, 1983). Consider the following example.

The personnel policies of an organization might indicate that performance appraisal is used to make decisions about pay—rewarding merit with salary increases. This claim may or may not be credible; an organization that claims to pay for performance but that allocates only a small amount of money for merit raises may be doing no more than giving lip service to actually rewarding performance (we consider this issue more fully in Chapter 13). However, even if the organization does make a good faith effort to reward good performance, there is no guarantee that individual supervisors and managers will adopt the goal of differentiating good from poor performance and allocating the largest rewards to the best performers. They may conclude, for a number of reasons, that it is more sensible to give everyone high ratings than it is to give high ratings to the best performers and low ratings to other employees who are not performing so well. In this chapter, we explore the reasons or rationale for this type of rating inflation.

Of course, rater goals are not the only factor that might lead to rating inflation. For example, Fried, Levi, Ben-David, and Tiegs (1999) showed that negative affectivity, the tendency to experience negative mood states over time and across situations, is related to rating inflation. They argue that raters who are high on negative affectivity will tend to pay attention to and recall more negative information, and as a result will arrive at more negative judgments about their subordinates. Given the strong situational pressure to give high ratings, raters whose initial judgments are more negative will be more likely to provide distorted ratings. Raters who are lower on negative affectivity might have more positive judgments in the first place, and have less need to inflate ratings.

Rater Goals

The specific goals pursued by raters will depend on a wide range of factors, including the values and experiences of the rater, the climate and culture of the organization, the performance level of subordinates, and so on.

A Taxonomy of Rater Goals

Murphy and Cleveland (1995) proposed that the most frequently pursued goals can be identified with one of four general categories: (a) task performance goals, (b) interpersonal goals, (c) self-serving goals,[1] and (d) internalized goals.

Task Performance Goals

If you ask raters why they give unrealistically high ratings, they frequently claim that they are using performance appraisal to increase ratees' performance levels or to maintain present performance levels. For example, giving an employee a favorable evaluation this year may lead to more rewards, more recognition, and perhaps a higher level of motivation to perform well in the future. Specific performance goals will necessarily depend on whether the ratee being evaluated is in fact (or in the rater's opinion) performing well. If the individual is performing well, the rater's goal might be to maintain that performance level. If the ratee is performing poorly, the rater's goal might be to increase motivation, or to identify and correct weaknesses.

A manager or supervisor who gives inflated ratings and overly positive feedback creates a variety of positive conditions that might contribute to improving performance in the future (or to maintaining current high levels of performance) in a number of ways. First, these ratings are likely to increase the probability that ratees will receive rewards and that they will view their supervisor and the performance appraisal system (and by extension, other HRM systems) favorably. This might make them more motivated and more receptive to feedback. On the other hand, ratings that are lower than ratees believe they deserve may lead to resentment, fewer rewards, and cynicism, and it is hard to see how any of these will contribute to future performance.

Interpersonal Goals

There is evidence that supervisors and managers use appraisal to maintain or improve interpersonal relations between the supervisor and subordinates. The interpersonal goal that has been discussed most frequently in the literature (e.g., Mohrman & Lawler, 1983; Napier & Latham, 1986) involves maintaining a positive climate in the work group. Because low ratings can lead to resentment and perceptions of inequity, the maintenance of a positive work group climate often involves rating inflation. We should note that some studies (e.g., Napier & Latham, 1986) have questioned the extent to which raters are concerned with interpersonal relations. If the rater doesn't care what others think of him or her, or of each other, this goal may not influence the rater's behavior. However, most raters do seem to care about their relationships with the people they evaluate, and if they don't they probably should. Positive relationships between supervisors and managers and employees can have a substantial effect on the ability and willingness of individuals and work groups to function well and perform effectively (Elangovan & Xie, 1999; Michael, 2014).

Sometimes, interpersonal goals may involve efforts to achieve or restore equity. If a supervisor believes that a specific subordinate deserved some reward that he or she did not receive (e.g., a promotion or raise), subsequent ratings may be distorted to achieve those outcomes. If work group members believe that rewards have not been distributed equitably, the supervisor might have a strong incentive to restore perceptions of equity. Yet another class of interpersonal goals are those that involve attempts to establish or maintain interpersonal influence. For example, a rater who demonstrates that he or she can have a substantial influence on administrative decisions (e.g., salary) may have a great deal of power and influence over subordinates.

Finally, and perhaps most significantly, supervisors and managers may be strongly motivated to avoid the sorts of interpersonal conflicts that can arise if employees receive lower ratings than they think they deserve. It is well known that supervisors dislike giving feedback (Cleveland, Lim, & Murphy, 2007), in part because they often expect that anything other than very positive feedback is going to be disregarded or is going to lead to arguments with or bad feelings on the part of subordinates. The motivation to avoid negative consequences can be substantially stronger than the motivation to pursue positive ones.

Self-Serving Goals

It seems clear that supervisors and managers can and do use performance appraisal to increase their standing and/or the work group's standing in the organization. There are several ways to accomplish this. First, uniformly high ratings might reflect positively on both the rater and the work group, especially if the rater is viewed as a credible source. Second, the standing of both the rater and the work group may be enhanced if work group members are given widely sought promotions and assignments; raters may distort their ratings to increase the likelihood of these valued outcomes. Third, the rater and the work group might both look better if the supervisor succeeds in transferring poor performers to someone else's workgroup. Again, performance ratings may influence this type of decision.

The ratings a supervisor or manager gives are also a reflection of his or her effectiveness. Consider two managers. Janice tells her superiors that all of her subordinates performed well last year. Frank tells his superiors that there were a few good performers, a few average ones, and a few real failures in his work group. One of the most important duties and roles of a manager is to create conditions that get the best performance out of the people he or she supervises, and by giving inflated ratings, Janice has just told her bosses that she was successful, whereas Frank (who might have given much more accurate ratings) has told his bosses that he failed. A rater who gives inflated ratings not only makes his or her subordinates look good, but also is engaging in self-promotion.

Although there are reasons to believe that many raters routinely give inflated ratings, this can become a self-defeating strategy. In particular, if too many raters use performance appraisal in blatantly self-serving ways, this is likely to lead to increased cynicism toward performance appraisal, perhaps even convincing organizations to curtail or abandon performance appraisal (Curtis, Harvey, & Ravden, 2005). If, as is true in most organizations, most raters turn in ratings that are moderately high, but perhaps not ridiculously so, rating inflation will end up being a matter of self-preservation rather than a way to make yourself look good. That is, the more extensive inflation is in an organization, the less benefit any individual rater is likely to get from giving inflated ratings and the more a rater who gives accurate ratings is likely to stand out from the pack. Unfortunately, in this instance, standing out means that a rater who gives accurate appraisals when virtually all of his or her colleagues give inflated ratings will probably ensure that his or her subordinates will look particularly bad in comparison with almost everyone else who is rated higher, will receive smaller raises and fewer promotions, and will probably resent the low ratings they receive.

Internalized Goals

Finally, it is worth considering goals that are the product of the rater's own values and beliefs. For example, a rater who believes that he or she should be honest in completing performance appraisals may turn in very different appraisals than a rater who believes that he or she should try to obtain valued rewards for subordinates. The rater's self-image may also affect rating behavior. Raters who believe that they are or should be participative leaders will probably devote more time and attention to feedback than will raters whose self-image is more authoritarian. Finally, raters might want to make sure that their subordinates do well in the organization, and might give high ratings because they know that good performance scores will be needed to qualify their subordinates for promotions or rewards (Smither, 2015).

The distinction between internalized goals and the other goals discussed in this section is that task performance, interpersonal, and self-serving goals are all instrumental in nature. That is, they involve giving high ratings in order to accomplish some particular end (i.e., increased performance, decreased conflict, self-promotion). Internalized goals are not directly instrumental, but rather represent rating strategies that reflect the rater's belief about the right thing to do. For example, a manager who is committed to giving his or her employees accurate feedback may do so even though he or she understands that accurate performance appraisals are likely to lead to negative outcomes, such as interpersonal conflict.

Alternative Goal Taxonomies

Wang et al. (2010) propose an alternate taxonomy of rater goals. First, they suggest that raters might pursue *harmony* goals, minimizing the distinctions between ratees. This might involve inflating the ratings of poorer performers, and perhaps forgoing inflation for better performers. Second, raters might pursue *fairness* goals, giving ratees the evaluations they deserve. Third, raters might pursue *identification* goals, with a focus on identifying strengths and weaknesses, and a relative de-emphasis on distinguishing between ratees. This might lead to smaller distinctions between ratees, but here the minimization of between-ratee differences would be a side-effect rather than the main goal of the rater. Finally, raters might pursue *motivational* goals, assigning the ratings they believe most likely to lead to increases in future performance. Wang et al. (2010) suggest that pursuit of these goals might lead to rating inflation for poor performers and rating *deflation* for good performers, but this is not the only pattern of rating that might be motivating. Depending on the rater's underlying assumptions about why subordinates perform well or poorly, ratings that are inflated, deflated, or accurate might be seen as motivating.

Development of Goals

When they first start their jobs, it is unlikely that supervisors or managers have well-articulated goals that direct their behavior in performance appraisal. Indeed, this group might be unique in that they function in a way that some appraisal researchers assume *all* raters function—their principal goal might be to do the best job they can to provide accurate and honest evaluations of their subordinates' performance. With more experience, raters are likely to develop a more complex, sophisticated set of goals that reflect the realities of their organization, department, or work group. In part, these goals will be shaped by their own experiences on the job. That is, they will learn what outcomes to expect as the result of giving good or poor ratings, and will also learn which of these outcomes is of concern to the organization. However, this type of trial-and-error process is not the only one involved in acquiring goals and beliefs regarding performance appraisal. Organizations make formal and informal efforts to socialize their members (for a wide-ranging review of research on organizational socialization, see Wanberg, 2012), and one aspect of being socialized into the role of "supervisor" or "manager" is to learn the appropriate behaviors and strategies for evaluating your subordinates.

Socialization Processes

Organizations are composed of a number of individuals who perform interlocking roles. Thus, it is important for newcomers to quickly learn the roles they are expected to carry out. Until they learn their roles, coworkers whose jobs are in some way linked to theirs will find it more difficult to perform their jobs. The precise content of what is learned in role socialization will vary from job to job, but there are some broad categories of learning that are likely to be present in most jobs (Fisher, 1986). First, and most obvious, you must learn how to do the job. Thus, training activities (both formal and informal) are one component of socialization. Second, you must learn the values and norms of the immediate work group. Third, you must learn the values, culture, goals, and norms of the organization; these are not always the same as those of the work group. Finally, you must incorporate your new role into your identity or self-image. One implication of this final category of learning is that a fully socialized individual will not only learn the role, but he or she will also internalize at least some aspects of the role.

The socialization process refers to the transition from the status of a naïve newcomer, who knows little about the role, to that of a knowledgeable insider, who understands the nuances of his or her role, as well as the relationship between that role and the roles of others in the organization (Chao, 1988). Much of this socialization occurs via the "role episode process," in which other members of the role set (i.e., other persons with a vested interest in how well you perform your role) send information evaluating the focal person's role performance (Graen, 1976). Behaviors that are consistent with the role are met with approval, while behaviors inconsistent with the role are sanctioned.

Table 12.1 Key Themes in Organizational Socialization

Socialization into an organization involves learning

- the basic *goals* of the organization;
- the preferred *means* by which these goals should be attained;
- the basic *responsibilities* of the member in the role, which is being granted to him by the organization;
- the *behavior patterns* that are required for effective performance in the role; and
- a *set of rules or principles* that pertain to the maintenance of the identity and integrity of the organization.

Schein (1967) suggests that organizational socialization is concerned with five general themes, summarized in Table 12.1. The second and fourth items in Table 12.1 are particularly relevant to performance appraisal. First, managers must learn the means by which key organizational goals are accomplished. Some of this learning takes the form of simple declarative and procedural knowledge—that is, learning who does what, how supplies and resources are obtained, who to go to for particular pieces of information, what forms to complete, how to submit reports, and so on. Some of this learning takes the form of *tacit knowledge* (Polanyi, 1966; Wagner, 1986)—that is, informally obtained knowledge that is rarely written down and is often informally obtained through experience. For example, you might learn through experience that the best way to get a task expedited is to demonstrate interest in or do small favors for the staff member who can best get that task moving.

Second, managers must learn what behavior patterns are most effective for performing their roles. One part of being socialized as a manager is to learn how people in your organization use performance appraisal, and how you should complete performance appraisals in ways that will maximize the likelihood of desirable outcomes and minimize the likelihood of undesirable ones. We will examine this hypothesis more fully in Chapters 13 and 14, but it is reasonable to believe that managers or supervisors with very little organizational experience (e.g., new hires, fresh from completing their BA or MBA) may approach performance appraisal in a relatively straightforward way, trying to do a diligent job observing the performance of their subordinates and giving them accurate ratings and feedback. Over time, they are likely to learn whether or not this strategy is an effective one, and also to learn how their rating strategy compares with the rating strategies of their peers. As they become more experienced as managers, they are likely to learn how to make the most effective use of performance appraisal as a tool for motivating employees, for building a positive work group climate, and even for self-promotion.

It is clear that some aspects of role socialization are easier than others. For example, expectations about the content of work are easily changed through experience (Wanous, 1980). Values and general work orientation are much less easily changed. In part this is because values are more firmly entrenched and more central than specific expectations about the content of work. However, there is another factor that influences the degree to which a person can be changed through socialization: the method of training. It might be difficult to impart values and norms directly; these may require somewhat indirect methods of socialization. However, individuals who share common experiences early in their tenure with the organization (e.g., while going through training and learning how to perform their jobs together) may also come to share perceptions and values that influence their subsequent behavior as raters. Weiss (1978), for example, suggests that organizational values are learned through modeling the behavior (and apparent values) of others. The ease with which modeling affects probably depends on the number of opportunities the focal person has to make relevant observations of others' behavior, as well as the power and status of the role sender (Fisher, 1986).

Organizational Culture and Rater Goals

Modeling of the behaviors of specific role senders is only one way in which newcomers acquire the values of the work group and the organization. These values and expectations that define the climate and culture of the organization influence many of the interactions between the focal person and other members of the organization (Rousseau, 1988; Schneider, Ehrhart, & Macey, 2011, 2013; Schneider & Reichers, 1983); these values will probably therefore influence the development of specific goals for performance appraisal.

Martin and Siehl (1983) note that the culture of an organization offers "an interpretation of an institution's history that members can use to decipher how they will be expected to behave in the future" (p. 52). Thus, the culture defines the behaviors that are either accepted or sanctioned, and the process of transmitting that culture (e.g., through stories or rituals) may serve as a powerful socializing force. In the context of performance appraisal, the culture might determine what approach to appraisal and what behaviors are evaluated positively or negatively in the organization.

It is reasonable to expect that beliefs, values, and norms regarding performance appraisal will often be a part of the organization's culture. In some organizations, appraisal is treated as important, and in others it is treated as a joke. Feedback is readily given and readily received in some organizations, and avoided in others (Sully, De Luque, & Sommer, 2000). Ratings are a critical determinant of decisions in some organizations but are ignored in others. It is possible that all of these aspects of appraisal are influenced by the culture of the organization, and the newcomers must learn that culture before they can use the appraisal system effectively to achieve their goals.

Organizational cultures can be described in terms of both the content and intensity of the values and beliefs that are part of that culture (Rousseau, 1988). The culture of an organization is most likely to affect the behavior of raters when: (a) it includes beliefs, values, and so on that are directly relevant to appraisal, and (b) those values, beliefs, and so on are strongly held. For example, Bjerke et al. (1987) studied the performance appraisal system used to evaluate officers in the U.S. Navy (the Fitness Report). They observed that there appears to be a strong norm in that organization that appraisals should be used to secure promotion for deserving subordinates (and for those who show promise of performing well in more responsible jobs). Most raters appeared to accept this norm, and filled out their appraisals with promotion in mind (Bjerke et al., 1987). These same conditions (i.e., strongly held beliefs about how appraisal should be done) may make it difficult to change the appraisal system in an organization without first changing the culture of the organization. However, it is possible that the relationship between performance appraisal norms and organizational culture is bidirectional. That is, changing the culture of an organization may make it easier to change performance appraisal systems, but it is also possible that making a substantial change in appraisal systems could change an organization's culture. In particular, an appraisal system that includes extensive opportunities for voice and for participation may help to make the entire organization more egalitarian and less rigidly hierarchical.

Goals as a Function of Subordinate Performance

To some extent, the rater's goals will be driven by the climate and culture that pervades the organization, as well as by the formal demands of the performance appraisal system. However, it is unlikely that a given rater pursues the same set of goals when evaluating each of his or her subordinates. Rather, there are likely to be very different goals when evaluating a chronically poor performer than when evaluating someone whose work is usually excellent. In the case of poor performers, the rater's goals might be to improve the performance, or to improve the chances that the worker will be transferred to some other work unit, or to justify a decision to discipline or dismiss the worker; the choice among goals such as these may depend in part on the extent to which the work group accepts your evaluation and supports your acting on it. In the case of good performers, the rater's goals will probably center on maintaining current performance; giving ratings that lead to raises, bonuses, or awards might be one mechanism for accomplishing this goal.

Although goals will be tailored, to some extent, to the subordinate, it is important to note that some goals are probably relevant to all subordinates (e.g., the goal of maintaining a positive work group climate), and probably do not vary as a function of the subordinate's performance level. The goals that are most likely to vary as a function of the ratee's performance are those in the categories we have labeled "task performance" and "self-seeking" goals.

Before discussing the specific goals that are pursued for good versus poor performers, it is useful to discuss a related, but somewhat more complex distinction among the individuals who report to a given supervisor. That is, the members of the work group can often be divided into two categories: (a) those in the in-group, and (b) those in the out-group. Raters probably pursue different goals when evaluating members of these two groups.

In-Group Versus Out-Group

Supervisors do not treat all subordinates the same way, nor should they. A widely researched theory of leadership, leader–member exchange (LMX) theory (Dansereau, Graen, & Haga, 1975; Erdogan & Liden, 2002; Gerstner & Day, 1997) suggests that different supervisor–subordinate dyads may involve distinctly different styles of interaction. In particular, the theory also suggests that dyads can be classified into those involving members of the in-group and those involving members of the out-group. Members of the in-group are chosen on the basis of (a) competence and skill, (b) the degree to which they can be trusted by the supervisor, and (c) their motivation to assume responsibility at work (Liden & Graen, 1980). In-group members perform the tasks that are most critical and challenging, whereas out-group members perform more mundane tasks. Furthermore, members of the in-group receive more information, as well as more confidence and concern from supervisors than do those in the out-group (Graen & Uhl-Bien, 1995).

One way to characterize the difference between the in-group and the out-group is that in-group members are treated by supervisors as trusted assistants, whereas out-group members are treated as hired hands (Vecchio & Gobdel, 1984). That is, in-group members are treated as valued colleagues; out-group members are treated as if they are temporary employees who are not really the concern of the supervisor. Supervisors typically devote more attention, resources, and the like to in-group members, even though out-group members might have a greater need for assistance, information, and other resources to perform their jobs.

Research on the LMX model suggests that supervisors adopt very different leadership styles when dealing with members of the in- versus the out-group (Graen & Uhl-Bien, 1995). In general, supervisors are more participative when dealing with in-group members, and more directive or authoritarian when dealing with out-group members. In the context of performance appraisal, this suggests that supervisors will more readily incorporate information from in-group members in their evaluations, and will usually devote more time and effort to feedback for in-group as opposed to out-group members.

Research on in- versus out-groups has dealt with several issues that are directly relevant to performance appraisal. First, and most fascinating, there is evidence that in-group status is related to subjective ratings of performance, but not to objective indices of performance (Vecchio & Gobdel, 1984). This implies that the perceived competence of in-group members may be an effect rather than the cause of their assignment to the in-group. It is possible that supervisors who feel that they can trust certain subordinates will perceive their performance as better than it actually is. A second relevant finding is that out-group members typically receive more extreme evaluations than do in-group members (Linville & Jones, 1980). These evaluations are usually negative, but the out-group member who actually does perform well may receive highly inflated ratings. One possible explanation for this effect is the fact that out-group members are expected to perform poorly. Performance that violates those expectations may attract more attention than performance that conforms to expectations (Hastie, 1980; Murphy, Balzer, Lockhart, & Eisenman, 1985).

Another finding in this research is that out-groups are seen as relatively homogeneous, whereas differences between in-group members are readily noticed by supervisors (Park & Rothbart, 1982). As a result, out-group members may all receive similar evaluations; stereotypes of the group as a whole may generalize to all group members. One implication of this finding is that performance evaluations may not provide feedback that is useful to out-group members. This feedback may reflect the group, and may not be tailored to specific individuals in the group. Another implication is that there will often be range restriction in ratings of out-group members. Thus, raters who have small in-groups and large out-groups may be especially likely to provide ratings that do not discriminate among most subordinates.

The terms in-group and out-group are not synonymous with “good performer” and “poor performer.” Nevertheless, research on in- versus out-group status may be useful in forming hypotheses about the goals that raters are likely to have when rating subordinates who have performed well or poorly. In particular, this literature suggests that good and poor performers may differ not only in the ratings they receive, but also in terms of how they are treated by supervisors in the appraisal process (e.g., the degree to which developmental feedback is given). In particular, there might be substantial differences in the goals being pursued by the rater when the ratee is a good, average, or poor performer.

Goals for Superior Performers

As we note in the next chapter, rating inflation is the norm in most organizations rather than the exception. One implication is that raters might find it difficult to give ratings to truly good performers that adequately reflect their actual performance level. If nearly everyone receives ratings at or near the top end of the scale, numerical ratings cannot be used to communicate to the ratee that he or she has really performed well. Thus, one goal that will often be relevant to the rater when evaluating a truly good performer will be to communicate to the ratee and the organization the fact that this individual’s performance is in fact better than that of others who have received high ratings.

As we noted earlier, raters may address this problem by developing a special vocabulary (Bjerke et al., 1987) or communication rules (Schall, 1983) that allow them to give ratings that are acceptable to all subordinates (in practice, this usually means giving inflated ratings to most ratees) and still communicate their evaluations of their subordinates. For example, a subordinate might receive a rating of “6” on a 7-point scale, and may be described as “somewhat above average” in the narrative or comment section of the appraisal. However, a phrase such as “somewhat above average” might connote poor performance; phrases such as “excellent in all respects of the job” might be used to signal performance that is actually above average. This process of developing special communication rules is especially likely to occur in organizations that have a strong and uniform culture. In such organizations, supervisors are probably well aware of what aspects of the job are treated as important, and are probably also aware of the special meanings that are given to specific evaluative terms in that organization.

A second goal that is highly relevant when evaluating good performers is to guide their development in such a way that they will be prepared for promotion, or for more complex assignments. The supervisor might accomplish this, for example, by giving feedback on aspects of performance that are not especially relevant for the subordinate's current job, but that are highly relevant for the next higher job in the organization. The popular literature contains many descriptions of mentoring and coaching: this approach to performance appraisal may correspond closely to one of the roles of a mentor or a coach.

One self-serving goal that might be pursued is to bring higher management's attention to the accomplishments of superior performers. Exceptional performance by your subordinates is likely to enhance your own standing as a supervisor; this is particularly true in organizations where the evaluation of the supervisor depends in part on the performance of the subordinates. We expect that the probability that this goal will affect rater behavior will be positively correlated with the subordinate's performance level. The better the subordinate's performance, the more the supervisor has to gain by bringing that performance to the attention of others in the organization.

Goals for Average Performers

It is reasonable to assume that in a large work group, the true distribution of job performance is reasonably normal (see, however, O'Boyle & Aguinis, 2012). One implication is that the number of truly good or truly poor performers will be small relative to the number of individuals whose performance can best be described as "average." Although average performers are the most numerous, we believe that raters do not pay a great deal of attention to their goals for evaluating average performers. There are several reasons for this. First, good performers are probably in the in-group and will therefore naturally receive much of the rater's attention and concern. Second, poor performers require attention; their low level of performance can directly affect the overall output of the work group, particularly if the work performed requires a high level of interdependence. Third, average performance is probably not as salient (or interesting) as either outstandingly good or outstandingly poor performance. As a result, average performers may not receive the attention they deserve.

There are a few goals that might become relevant for average performers, depending on the characteristics of both the person and the situation. First, there might be some individuals who because of their trustworthiness and responsibility, would be good candidates for the in-group. The rater's goal when evaluating these individuals might be to improve their performance, and bring it up to the level of other in-group members. Second, the rater might have a strong incentive to help improve the performance of individuals who are members of specific groups. For example, the supervisor whose values encourage affirmative action efforts might devote considerable time and attention to the development of members of protected groups whose performance is at or below the work group's average. On the other hand, the supervisor who is prejudiced against members of specific groups might be inclined to give them harsh ratings; the underlying goal may be to remove these individuals from the work group.

Goals for Poor Performers

The research literature dealing with supervisors' reactions to poor performance is extensive. Although relatively few studies deal directly with performance appraisal, several conclusions can be drawn about the appraisal goals that might be pursued when evaluating poor performers.

It is clear that managers prefer to avoid dealing with the problem of poor performance (Mitchell & O'Reilly, 1983). In the context of appraisal, raters might avoid the problem by giving relatively high ratings, accompanied by vague and noncommittal comments. In this case, feedback is likely to be perfunctory, and little effort will be expended to improve the subordinate's performance. In many cases, however, it may be impossible to ignore poor performance, either because it is extremely bad or because there are situational factors (e.g., coworkers' resentment) that force the rater to attend to the problem of poor performance.

The rater's reaction to a poor-performing subordinate probably depends on the attributions he or she makes when deciding about the cause of that performance (Jones & Nisbett, 1971; Mitchell & O'Reilly, 1983; Ross, 1977). Mitchell, Green, and Wood (1981) described several different categories of attribution, including those that explain poor performance in terms of a lack of ability, a lack of values and standards, an unfavorable environment, or a variety of temporary barriers to performance (e.g., family crises). These specific categories of attributions can in turn be classified in terms of two dimensions that are frequently encountered in attributional research: (a) attributions to internal versus external causes, and (b) attributions to stable versus temporary causes. The supervisor's response may be very different, depending on the type of attribution.

Attributions depend on a number of factors, including the behavior itself, and the values, standards, and inclinations of the supervisor (e.g., some supervisors make more internal attributions than others). Another factor that appears to influence the supervisor's attributions is the outcome of the behavior. The same behavior might be evaluated differently, and different causes might be cited for that behavior, depending on the seriousness of the outcomes of that behavior. For example, if two subordinates follow the exact same unsafe work practices, and one of them causes or experiences an accident, they will probably be evaluated differently, with harsher evaluations for the person involved in the accident.

Research on attributions suggests systematic differences in the supervisors' versus subordinates' explanations for poor performance. In general, supervisors make more internal attributions (e.g., the person did not perform well because of poor work values), whereas subordinates are biased in favor of external attributions (e.g., constraints in the situation made it impossible to perform well; Gioia & Sims, 1985; Ilgen, Mitchell, & Frederickson, 1981). The tendency of supervisors to make internal attributions increases the likelihood that their responses will involve discipline or punishment for the poor-performing subordinate.

Arvey and Jones (1985) note that the topic of discipline or punishment in organizational settings has been largely ignored by behavioral scientists. Studies that have examined punishment suggest that performance appraisal is sometimes used as a vehicle for administering discipline. For example, Jones, Tait, and Butler (1983) studied the degree to which different supervisory behaviors were perceived by subordinates as punishing. Behaviors that were cited as being very punishing included: (a) telling a superior about your mistakes, (b) documenting the negative things you do, and (c) publicly praising everyone in the group but you. These and other punishing behaviors may be used differently in in-groups versus out-groups (Arvey & Jones, 1985); in general, the frequency and severity of punishment will be lower in the in-group.

If the attribution is external, the supervisor may avoid punishing the subordinate. Task performance goals for poor-performing subordinates (assuming that the attributions are external) will probably involve efforts to develop the subordinate, to maintain or increase motivation, and to move the subordinate to circumstances that are more favorable (e.g., a less difficult assignment or a different, more supportive work group). The rater who intends to develop a poor-performing subordinate will probably (a) give higher ratings than deserved, in an effort to avoid organizational sanctions and maintain motivation, (b) devote more attention to feedback, and (c) identify weaknesses in feedback sessions, but not in the performance evaluation forms that are turned in to the organization.

When attributions are internal, strategic goals that involve getting rid of the poor-performing subordinate may strongly influence rating behavior. This will depend in part on how bad the performance really is. The worse the performance, the stronger the incentive will be to get rid of the subordinate. The strength of this goal may also depend on what specific aspects of performance are below standard. For example, supervisors might be more strongly inclined to get rid of subordinates whose poor performance reflects an inability to get along with coworkers than those whose performance reflects sincere effort but low ability. In general, performance deficiencies that may affect other coworkers may present a stronger incentive to get rid of the subordinate than will deficiencies that involve the individual only.

Raters' responses to poor performance may depend in part on their level of power in comparison to the person they are rating. Raters who have substantial power are more likely to make a proactive response, either engaging the ratee in activities designed to develop better performer or confronting the ratee over his or her shortcomings. Ratees who have less power may be somewhat more likely to avoid interacting with poor performers or to compensate for their poor performance in some way (Ferguson, Ormiston, & Moon, 2010; see, however, Liden et al., 1999).

The exact rating strategy for accomplishing the goals described above will depend in part on the policies of the organization. For example, if the organization is very reluctant to fire poor performers, it may not be a good idea to give low ratings. If the subordinate consistently receives low ratings, it may be very difficult to persuade the organization to transfer him or her to another work group. The attractiveness of "dumping" a poor-performing subordinate on another work group will vary as a function of the nature and size of the organization. In a small organization, supervisors are more likely to know one another, and to be friends. It may be very difficult, in this case, to transfer a poor-performing subordinate to someone else's work group without negatively affecting your relationship with that supervisor.

Analyzing Performance Rating Strategies as a Choice Between Alternatives

A series of influential articles in the 1980s (e.g., Feldman, 1981; Landy & Farr, 1980) encouraged performance appraisal researchers and practitioners to think about performance appraisal as an information processing exercise. That is, topics such as attention, memory, and the integration of information were held out as important for making progress in understanding performance rating and performance evaluation. This type of research did make many contributions, but by the 1990s the field was shifting its attention from cognition to understanding how the context in which performance appraisal occurs drives rating behavior and toward examining the factors that motivate raters to give high or low ratings (Banks & Murphy, 1985; Longenecker et al., 1987; Levy & Williams, 2004; Murphy & Cleveland, 1991, 1995; Murphy & DeNisi, 2008).

As we have noted in several places, it is important to distinguish between the judgments raters form when completing performance appraisal forms and the actual ratings they assign (Landy & Farr, 1980; Murphy & Cleveland, 1995; Murphy & DeNisi, 2008). Judgments about performance are not something that occur only when it is time to complete performance appraisal forms, but rather they represent an ongoing process that involves a number of separate cognitive and affective processes (e.g., what did you observe, how did you interpret it, how do you feel about the person being evaluated?).

Regardless of the judgments they may have formed about the performance and effectiveness of their subordinates, raters need to make two choices when it comes time to complete performance appraisal forms. First, should the rater record the number that best represents his or her judgment about the performance of the ratee? There may be a number of good reasons to record a number that is not identical to the rater's judgments about the ratee. If the rater does choose to distort (i.e., put a number on the appraisal form that is different from his or her private judgment), he or she will have to choose whether to give higher ratings or lower ratings than the ratee appears to deserve. Schematically, the choice raters face when completing a performance appraisal form is portrayed in Figure 12.1.

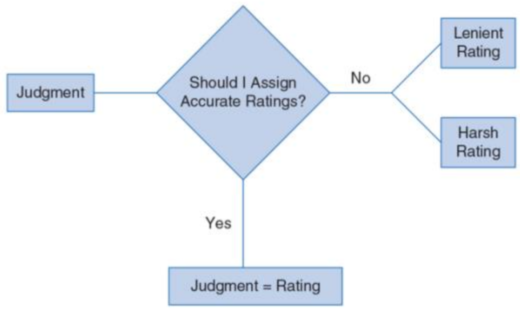

Figure 12.1 From Judgments to Ratings

If performance appraisal is approached from the perspective of a cost-benefit analysis, it is clear that supervisors have few good reasons to give accurate ratings and even fewer to give harsh ratings (i.e., ratings that are lower than the ratee's performance would warrant), but that they have many good reasons to give inflated ones (Longenecker et al., 1987; Murphy & Cleveland, 1995). Mohrman and Lawler (1983) note that there are typically few rewards and many penalties for doing accurate appraisals. Various studies have shown that raters manipulate performance ratings to avoid negative consequences (e.g., Bernardin & Beatty, 1984; Cleveland & Murphy, 1992; Longenecker et al., 1987; Spence & Keeping, 2010; Tziner, Murphy, & Cleveland, 2005), to comply with organizational norms (Bernardin & Beatty, 1984; DeCotiis & Petit, 1978; Spence & Keeping 2010; Tziner et al., 2005) and to further their own interests (Spence & Keeping, 2010; Villanova & Bernardin, 1989).

A number of authors (e.g., Kane, 1994; Murphy & Cleveland, 1991, 1995) have proposed models of raters' decisions to distort ratings. These models start by considering the likely outcomes if raters give performance ratings that are either accurate or inflated.[2] Because most employees overestimate their own level of performance, performance ratings that are reasonably accurate will often seem unfair and harsh, whereas ratings that are inflated will seem reasonable and fair. The models used to describe and explain distortion in performance ratings can give you important insights into why performance ratings often seem to fail in organizations.

Outcomes of Accurate Versus Distorted Ratings

Suppose you are a typical employee and you think your performance is above average. You would be in good company—most people rate themselves above average on a wide range of indices. For example, about 80% of all drivers think they are above average in driving skills (McCormick, Walkey, & Green, 1986). Accurate ratings of your job performance will probably be lower than what you expect and believe you deserve, and this is likely to have several consequences. First, if you expect positive feedback but receive feedback that your performance is not as good as you thought, you will probably feel disappointed and may even feel you have been treated unfairly. Your relationship with your supervisor may deteriorate. You might not receive as big a raise as you think you deserve, and you may end up concluding that the whole performance rating system is unreliable. On the other hand, if your supervisor gives you ratings that are higher than you actually deserve, you will probably be more satisfied and will probably have a more positive feeling about performance appraisal. It is perhaps ironic, but there is evidence that distorting ratings may actually *contribute* to the perception that they are accurate and fair, and giving fair and accurate ratings can contribute to the perception that ratings are rigged.

In general, we can group the probable outcomes of giving accurate versus inflated ratings into four categories. First, there are interpersonal consequences. Giving your subordinates ratings that they perceive as unfairly low will probably not be good for the future relationship between you and that subordinate. Second, there are systemic consequences. As we noted in Chapters 9 and 10, the ratings people receive influence their reactions to and their beliefs about the performance appraisal system in their organization. Receiving ratings that you believe are unfairly low could shake your confidence in this system. Third, there are practical consequences.

Ratees who receive accurate ratings may miss out on chances for raises and promotions if their peers are receiving inflated ratings. Finally, there are self-serving consequences. A manager who tells the organization that all of his or her subordinates are performing well is also indirectly telling the organization that he or she must be an effective manager (Bowman, 1999; Murphy & Cleveland, 1995), whereas a rater who gives accurate (i.e., lower) ratings is telling his or her superiors that he or she is not doing the thing managers are most responsible for—that is, helping their subordinates to perform their jobs at peak efficiency. Table 12.2 lists a number of the likely consequences of giving accurate versus inflated ratings.

Table 12.2 Probable Consequences of Giving Accurate Versus Inflated Ratings

Accurate
Interpersonal Consequences
Disappointment and resentment on the part of the ratee
Deterioration in relationships with raters
Systemic Consequences
Reduced confidence in the rater
Reduced confidence in performance appraisal
Practical Consequences
Fewer rewards
Decreased motivation
Self-Serving Consequence
Negative impressions on the part of your own superiors
Inflated
Interpersonal Consequence
Positive relations with subordinates are maintained and reinforced
Systemic Consequences
Ratees will perceive performance evaluations as fair and accurate
More trust in and commitment to the organization
Practical Consequences
Ratees receive valued rewards
Motivation increases

Self-Serving Consequence

Positive impressions are conveyed to your own superiors

Contingencies

In addition to thinking about the possible consequences of giving accurate versus inflated ratings, raters may need to consider the strength of the link between ratings and consequences. For example, if a supervisor gives inflated ratings, this increases the likelihood of subordinates getting raises and promotions, but there is no guarantee that they *will* receive these raises or promotion. If other raters also give inflated ratings, simply inflating your own ratings may not be enough to guarantee that your subordinates get raises. On the other hand, if other raters are lenient and you choose to give accurate ratings, this might have a strong and immediate effect on outcomes that are important to your subordinates. Similarly, giving high ratings *may* lead to good feelings on the part of your employees (however, they are likely to believe that they fully deserve good ratings, so it is not a good idea to expect gratitude), but you can be pretty sure that if you give accurate ratings, which appear to be unduly harsh to your subordinates, they will resent what appears to be shabby treatment.

On the whole, the links between ratings, behaviors, and negative outcomes are likely to be stronger than the links between ratings and more positive outcomes. That is, giving inflated ratings will often help, but since most other raters are also giving inflated ratings, they may not have a strong effect. On the other hand, if most other raters give inflated ratings and you give accurate ones, these low scores will really stand out, and they will be likely to trigger some unfortunate consequences.

Numerous models of decision making and choice start with the two variables described here—the likely outcomes of different choices and the perceived links between the choices a rater makes and those outcomes—and combine these two factors to make predictions about what choices raters will make (Beach, 1990; Busemeyer & Townsend, 1993; Edwards, 1980; Edwards & Newman, 1982; Hammond, McClelland, & Mumpower, 1980; Kahneman, 2011; Kahneman & Tversky, 1979; Keeney & Raiffa, 1976; Roe, Busemeyer, & Townsend, 2001; Vancouver, Weinhardt, & Schmidt, 2010). These models differ in a number of important ways, but they all agree that people are likely to choose a particular course of action if: (1) this action is linked to valued outcomes, and (2) the links between that choice and those outcomes are strong and dependable. Thus, a choice that might lead to highly valued outcomes might not look so attractive if the links between that choice and those outcomes are uncertain or weak. For example, a person might believe that the chance to go into space would be a wonderful accomplishment, but also understand that the likelihood of being selected for and succeeding in astronaut school is extremely low. Virtually all models of decision making and choice suggest that even though the outcome might be a very desirable one, if the perceived likelihood of accomplishing that outcome is low, there will not be much motivation to pursue that choice of action.

Integrating Multiple Perspectives on Motivation to Distort Ratings

Murphy and Cleveland’s (1995) discussion of rater motivation focuses largely on the rater’s decision of what rating to record, and is based on the assumption that this rating might not always correspond to the rater’s judgments about ratee performance. Like many other models (e.g., Ballard et al., 2016; Spence & Keeping, 2011), their focus is largely on conscious distortion in rating—that is, the decision at the time performance ratings are required to give employees ratings that are higher than they truly deserve. Other authors suggest that the processes that lead to widespread rating inflation occur at a very different point in time and unfold in ways that models of conscious distortion fail to capture.

Spotlight 12.1 A Multiple Goal Pursuit Model

Ballard, Yeo, Loft, Vancouver, and Neal (2016) have developed and tested an integrative model of motivation and decision making, referred to as the MGPM* model. This model is an extension of an earlier model developed by Vancouver, Weinhardt, and Schmidt (2010), and it is a very useful framework for understanding how raters might make choices among various rating strategies when evaluating the performance of their subordinates.

Spotlight 12.1 A Multiple Goal Pursuit Model

Ballard et al. (2016) start by noting, "people often pursue multiple, competing goals, striving to achieve desired outcomes while avoiding undesirable outcomes. Because individuals have limited time and resources, they often have to choose which goal to prioritize at the expense of progress toward others" (p. 1240). They developed a model to help them study and predict the choice among different courses of action. Some of the key concepts of this model are summarized in Table 12.3.

One of the key insights in the MGPM* model is that the choice to pursue particular goals is not a "one-and-done" event. That is, it is wrong to assume that individuals choose one goal, then continue to pursue it to the exclusion of other goals over time. Rather, this model assumes that the attractiveness of different goals can change over time, especially as people make progress toward accomplishing a goal, and people might switch goals in a predictable way. The other key insight of this model is that framing matters. The value that is attached to a goal might change depending on whether it is framed and understood by the decision maker to be an approach goal (e.g., maintain good relationships with your subordinates) or an avoidance goal (e.g., avoid interpersonal conflicts over performance evaluation). This theory helps to explain why people often shift priority from approach goals to avoidance goals over time.

Table 12.3 Key Concepts in the MGPM* Model

- Discrepancy—difference between the current state of a variable and the reference state, or the goal
- Expected utility—multiplicative function of the value of a goal and the perceived likelihood of reaching that goal
- Goal types—approach goals involve a desirable state people want to pursue, whereas avoidance goals involve undesirable states people want to avoid
- Dynamism—the value of a goal can change as a function of the size of the discrepancy; the motivational value of approach goals decreases the closer you are to accomplishing the goal, whereas the motivational value of avoidance goals increases the closer you are to the undesirable state
- Attention weight—likelihood that people will attend to each possible consequence of attaining a goal

According to the MGPM* model, raters who have the option to pursue a number of different goals when completing performance appraisals make this choice by considering the possible consequences of achieving different goals, the likelihood that they *can* accomplish these goals, and the value of the different consequences. Their assessments of several of these factors might change as the discrepancy between the current state and the desired goal state grows or shrinks. Thus, we might assume that the goals pursued by raters when they complete performance appraisals might differ both across raters (Murphy et al., 2004) and within raters across time.

Finally, this model suggests that the motivational power of particular goals change as you approach accomplishing each goal. On the whole, the power of positive goals (e.g., goal of seeking a reward) to motivate behavior *decreases* as you approach accomplishing the goal, which implies that the anticipation of good outcomes might be more motivating than the outcomes themselves. On the other hand, the motivating power of negative goals (e.g., goal of avoiding punishments) *increases* the closer you get to reaching them. Thus, raters might not worry much about negative outcomes that might occur in the distant future, but might be strongly motivated to avoid imminent threats. If we apply this to performance appraisal, the MGPM* model suggests that raters might be influenced by potential positive outcomes as they are thinking about how they want to approach performance appraisal in the future, but as they approach the time when they will actually have to prepare and submit performance ratings, they become more and more focused on avoiding negative consequences.

Ballard et al. (2016) suggest that one practical application of their theory is to illustrate the potential benefits of reframing goals in ways that allow people to appropriately balance competing objectives over time. For example, using performance feedback with the goal of building and maintaining good relationships with subordinates might entail different choices over time than using performance with the goal of minimizing conflict with employees. An approach goal might encourage more frequent positive feedback (consistent with the philosophy of performance management), whereas an avoidance goal might encourage raters to give inflated ratings when they complete performance appraisals in an effort to avoid conflict in the annual appraisal meeting.

Harris (1994) notes that the same factors that motivate raters to distort the ratings they provide (e.g., rewards, avoidance of negative consequences, impression management) might also influence rater behaviors throughout the process of acquiring information about ratee performance, recalling and integrating this information to form a judgment, deciding what rating to assign, then deciding what feedback to provide. Although Harris recognizes the potential role of rating distortion, he makes the important point that during the time in which raters complete performance appraisal forms, this is probably not the only circumstance in the appraisal process in which rater motivation matters. Rather, raters who are motivated to provide accurate ratings are likely to be more careful in their observations, to take steps to accurately recall ratee behavior (e.g., keeping behavior diaries), and might be more mindful when pulling this information together to reach judgments about ratee performance. Raters who know that at the end of the day they are going to give high ratings to everyone will probably put less time and effort into evaluating their subordinates.

So far, we have focused on understanding what motivates raters to distort their ratings but Spence and Keeping (2011) note that the same forces that often lead to distorted ratings can in some circumstances provide motivation to rate accurately, especially if accurate rating serves at least one of three goals: (1) enhancing relationships with subordinates, (2) improving the rater's image within the organization, or (3) increasing the likelihood of achieving the goals of the organization. They argue that some employees want accurate feedback, and that they may value accurate ratings because these ratings can help them improve their future performance. Similarly, in an appraisal system where just about everyone gives inflated ratings, that small group of raters whose evaluations are more accurate and honest might be recognized as people who are doing their best to make the appraisal system work as intended. Finally, one of the cornerstones of performance management is the idea that feedback about performance can be used to direct employee behavior in ways that will maximize the likelihood of accomplishing important organizational goals, and accurate feedback (even if that feedback is painful) may be more useful than distorted feedback.

Spencer and Keeping's (2011) point that there can be strong motives to provide accurate rather than distorted ratings is well taken. Other models (e.g., Murphy & Cleveland, 1995) have focused on raters' motivation to distort ratings, but have not given much consideration to raters' motivation to provide accurate ratings. What both Spencer and Keeping (2011) and Murphy and Cleveland (1995) failed to explicitly consider is the possibility that accuracy in rating might help in achieving some goals (e.g., advancing the interests of the organization) while decreasing the likelihood of other goals (e.g., enhancing relationships with subordinates). The question then is how raters will balance the potential gains and losses associated with accuracy in rating.

There is little empirical research on how these tradeoffs might be resolved, but we can suggest some possibilities. Distortion in rating (usually leniency) is likely to help with two things. First, it helps in maintaining a positive relationship with subordinates. Second, it helps with maintaining the image of success as a supervisor or manager (because all of your subordinates are "successful"). On the other hand, accuracy is more likely to help the organization achieve its goals. It seems likely that raters will pick that which is good for the organization over that which is good for themselves and their subordinates if (1) there is a strong identification with the goals of the organization, (2) the rater is sufficiently confident of his or her image in the organization that the incentive to enhance that reputation via rating distortion is minimal, and/or (3) relationships with subordinates are either so good that lower ratings will not hurt them, so negative that lenient ratings will not help, or so unimportant that they do not really matter to the rater. We believe that raters who either (a) depend very much on the cooperation and good will of their subordinates or (b) have not established a strong and lasting image in the organization will be most likely to distort their ratings.

Avoiding Negative Outcomes

One of the themes of numerous models of choice and decision making is that negative outcomes often have stronger influence on decisions and choices than positive ones. For example, according to prospect theory (Kahneman & Tversky, 1979), losses are often seen as more substantial than equivalent gains. That is, gaining $200 feels good but losing $200 feels *very* bad, and people will work harder to avoid a $200 loss than they will to pursue a $200 gain.

The effect of negative outcomes is likely to be especially important in performance appraisal because of the certainty of positive and negative outcomes if inflated versus accurate ratings are given. That is, in a context where most raters give and most ratees expect highly positive ratings, the likelihood of resentment, reduced rewards, conflict, and the like may be pretty high if you give accurate ratings, whereas giving inflated ratings does not really do anything to make you stand out from the crowd. That is, raters have a number of reasons to give inflated ratings, but they have even stronger reasons to avoid giving accurate ratings. Thus, in an organization where accurate performance feedback really is desired, it might be more important to find ways to shield raters from negative consequences than it is to give them positive reasons to provide accurate ratings. We take up this suggestion in more detail in Chapters 13 and 14.

Explaining Rating Distortion

So far, we have focused on motivational models that use some combination of beliefs about the outcomes of different choices when completing performance ratings and about the strength of the links between ratings and those outcomes to make predictions about whether raters will give accurate or distorted ratings. These models have considerable explanatory power, but they probably put too much emphasis on cold, rational choice. It is misleading to assume that raters simply weigh the consequences of different choices they might make when rating and blindly pursue the course of action with the highest expected utility (that is, value of outcomes multiplied by the perceived link between particular choices and these outcomes). There are attributes of individuals and of the context in which ratings occur that may lead raters to choose evaluation strategies that may not be optimal from a strictly rational point of view.

Individual Differences

Our discussion of leniency in rating has focused mainly on motivational factors (e.g., a desire to avoid negative relationships with subordinates), but there are also individual differences that influence the likelihood that raters will be lenient in their ratings. For example, Markus and Kitayama (1991) proposed that some supervisors have an *independent* view of human behavior, with an emphasis on the role of the individual and his or her own characteristics in determining behavior, whereas others have an *interdependent* view of human behavior, with a stronger emphasis on the role of society and the context in shaping behavior. There is evidence that raters whose worldview emphasizes interdependence tend to give higher performance ratings than raters who emphasize independence (Mishra & Roch, 2013).

The independent–interdependent dichotomy suggested by Markus and Kitayama (1991) is clearly related to the broader cultural dimensions of collectivism–individualism (Hofstede, 2001). There is evidence that raters from more collectivistic cultures tend to be more lenient (Ng et al., 2011). This same study also presented evidence that the cultural dimension of Power Distance may help to explain the tendency of some rating sources (e.g., peers) to be more lenient than others.

In addition to broad worldviews and cultural factors, there are aspects of raters' personalities that can influence performance ratings, although the effects of personality on leniency are not always clear. For example, the weight raters place on different performance dimensions when forming their overall evaluations can differ, depending on rater personality. Ogunfowora, Bourdage, and Lee (2010) reported that raters who are high on Openness to Experience tend to place more emphasis on success in dealing with complex and unpredictable environments, whereas raters who are high on Modesty tend to place more weight on maintaining control and abiding by rules. Although there have not been direct empirical tests of the links between these personality dimensions and rating inflation, it is reasonable to expect that raters who are high on Agreeableness will be more likely to be lenient than less Agreeable raters.

Finally, there are a range of beliefs about the task of performance appraisal and about the organization that can influence distortion in rating. Rater's beliefs that they have the tools, information, and skills to evaluate their subordinates' behavior are an important component of the motivation to rate accurately (Bernardin & Villanova, 2005; Brutus, Fletcher, & Baldry, 2009); there is even evidence that perceptions of self-efficacy as a rater are related to the quality of rating data (Bernardin & Villanova, 2005; Tziner & Murphy, 1999). Tziner (1999) and Tziner and Murphy (1999) present evidence that beliefs about the organization (e.g., perceptions of the climate of the organization, commitment to the organization), relationships between supervisors and subordinates, and perceptions of self-efficacy as a rater (i.e., beliefs that you have the necessary information and skills to accurately evaluate rate performance) all influence the likelihood that raters will inflate their evaluations of subordinates.

Contextual Factors

In several previous chapters, we have discussed the effects of contextual factors on performance appraisal processes and outcomes. The term *contextual factors* can include a wide range of variables, but there is one particular aspect of the rating context that appears to be particularly important in influencing rating distortion: the way organizations react when raters submit performance ratings that appear to be distorted. One way of determining whether organizations *actually* care about performance appraisal is to ask whether they attempt to reward raters for accuracy or to sanction raters for providing ratings that are clearly inaccurate. If there are genuine rewards for providing accurate ratings and useful feedback, or genuine sanctions for providing inaccurate evaluations, this might decrease the likelihood of rating distortion.

Bjerke et al.'s (1987) study of the U.S. Navy's Fitness Report (the form used to evaluate the performance of Naval officers) suggested that there were usually few negative consequences for raters who give inaccurate, inflated ratings (i.e., raters are not penalized for inaccuracy). However, they also found that there were limits to the amount of rating inflation that was viewed as acceptable, and that in some cases, raters were criticized for overdoing it—that is, providing ratings that were so blatantly unrealistic as to violate organizational norms. On the whole, we suspect it is easier to sanction rating distortion when it is so blatant that it offends the sensibilities of other managers and executives who are also providing inflated ratings than it is to reward accurate ratings. In virtually any organization, norms develop regarding the extent to which employees and managers are allowed to bend rules and skate around the edges of personnel policies versus doing things strictly by the book.[3] Raters who go beyond these implicit bounds when they rate their subordinates may be sanctioned.

Rewarding accuracy is arguably more difficult because it may be more difficult to tell if a particular rater is being accurate than it is to tell if he or she is distorting. A rater who consistently tells you that all of his or her subordinates are excellent in all aspects of performance all of the time is almost certainly pulling your leg. Ratings like this might serve a number of goals that are important to the rater, but it should be obvious that they are likely to be distorted and untrustworthy. On the other hand, a rater who tells you that some subordinates are performing well, some are performing poorly, and some are near average *might* be giving you accurate information, but it is also possible that the distribution is realistic but that the particulars are all wrong. That is, the rater might be giving you inaccurate information about *who* is a good performer, a poor performer, or an average performer, and if you do not have other information that can be used to cross-check these ratings, it might not be possible to tell if that rater really is providing accurate evaluations.

Summary

The best explanation for rating inflation, which is rampant in most organizations, is that raters believe that giving high ratings (even unrealistically high ratings) will lead to valued outcomes and that giving accurate ratings will lead to undesirable outcomes. An analysis of the goals that appear to influence raters' choice to distort ratings suggests that distortion can influence task performance, interpersonal relationships, and the rater's own standing in the organization. Raters appear to develop their goals and strategies for using appraisal to accomplish these goals over time, and organizations often employ a mix of formal and informal socialization processes that are likely to help shape these goals and strategies.

Models of decision making and choice provide useful insights for understanding decisions to provide either accurate or distorted ratings. In particular, these models suggest that we need to consider both the range of outcomes that are likely to be tied to accurate versus inaccurate ratings and to the strength of the link between the ratings and these outcomes, and that the choice to distort ratings is often driven by an understanding that accurate ratings are very likely to lead to negative outcomes and that distorted ratings increase the probability of several positive outcomes. However, goals are not the only factor influencing rating distortion. There are both individual differences and a number of features of the context in which rating occurs that often push raters to give high ratings to virtually all of the people they rate.

Exercise: Build a Goal Assessment Tool

In this chapter, we have argued that raters are often motivated to inflate performance ratings in order to avoid conflict, to increase the motivation of their subordinates, to make themselves look good, and so on. You might get insight into rating dynamics if you knew what goals individual raters were pursuing, and the relative emphasis they placed on different goals. However, we suspect that if you ask supervisors "are you giving your subordinated high performance ratings in order to make yourself look good?", you may not get an honest answer. A better approach might be to assess whether: (a) ratings are likely to be inflated in your organization, and (b) conditions in your organization are likely to create pressures to inflate performance ratings.

First, take a careful look at both the distribution of performance ratings and the relationships between performance ratings and other likely indicators of performance or success. Suppose, for example, that you find that performance ratings are very strongly concentrated in the top few rating categories. This would be *potential* evidence of rating inflation, but suppose you also find that the people who receive the highest ratings are consistently more productive, or achieve better outcomes than the people who receive lower ratings. High average ratings, by themselves, do not allow you to draw strong conclusions about rating distortion, and if other evidence supports the usefulness of ratings (albeit with restricted range), you might conclude that whatever inflation is present is not necessarily something to be concerned about.

Suppose you determine that ratings do appear to be seriously distorted. The taxonomies of rater goals presented earlier in this chapter suggest that there are several likely explanations for ratings inflation. First, raters might inflate ratings to improve their relationships with subordinates. Second, raters may inflate ratings to help their subordinates obtain valued outcomes (e.g., raises). Third, raters might inflate ratings to make themselves look good. Organizational surveys might allow you to assess the likely strength of each of these motives.

First, you might survey employees to evaluate their perceptions of the fairness and accuracy of the ratings they receive, with a particular focus on the extent to which employees believe *their supervisor* is giving them fair evaluations. The more sensitivity to the fairness and accuracy or ratings from direct supervisors, the more pressure there is likely to be on raters to inflate ratings. Second, you can survey higher-level managers to determine how they evaluate the ratings supervisors provide. In organizations where raters are viewed unfavorably when their subordinates receive less positive evaluations, and viewed favorably when their subordinates are rated as effective performers, supervisors are more likely to pursue self-serving goals. Third, you can assess both the actual and the perceived links between performance ratings and valued rewards in your organization. If supervisors and employees believe that high performance ratings are a key to getting raises and promotions, this is likely to lead to motivation to inflate ratings—even if these beliefs are incorrect. As we will note in Chapters 13 and 14, many organizations claim that pay is linked to performance, but put in place policies (e.g., small merit pay pools) that make these links weak and inconsistent. Employees and supervisors, however, sometimes do not have the information to make informed judgments about the actual strength of these links, and they are more likely to act on their perceptions and beliefs.

Suppose you find that the conditions in your organization are ripe for rating inflation (e.g., employees believe that there are strong links between ratings and rewards and are strongly concerned that they are being unfairly evaluated, ratees are evaluated more positively if they hand in high ratings for their subordinates). It is not always clear that these conditions can, or even should be changed in organizations—the pursuit of more accurate performance ratings is valuable only if organizations actually use this information to make better decisions and to improve their workforce. The sort of assessment envisioned here will at least give you better insight into why ratings might be inflated and what might be done to counter that inflation, but the decision about whether or not these changes *should* be made will not always be a simple one.

Notes

1. Murphy and Cleveland (1995) labeled these “strategic goals,” but the label “self-serving goals” is more precise, because the strategic objective here is to make the rater look good.

2. Raters who distort their ratings typically give their subordinates more positive evaluations than they deserve, but there are instances where purposely harsh ratings are assigned to send a message to or to punish an ineffective employee.

3. If the norms of an organization, or a work group, are highly tolerant of bending and perhaps even breaking rules, the likelihood of counterproductive behaviors will probably increase (Murphy, 1993).