The Influence of a Manager’s Own Performance Appraisal on the Evaluation of Others

Gary P. Latham*, Marie-Hélène Budworth**, Basak Yanar* and Glen Whyte*

*Joseph L. Rotman School of Management, University of Toronto, Toronto, ON, Canada M5S 3E6.
latham@rotman.utoronto.ca
**School of Administrative Studies, York University, Toronto, ON, Canada M3J 1P3

This study examined the possibility that the performance appraisal process is affected by a pervasive and inherent effect that has heretofore been unidentified. This effect derives from the results of the performance appraisal most recently performed on the manager who subsequently conducts appraisals of others. The nature of this effect was examined in four studies. In a case study, the ratings received by two area coordinators in a university academic department affected their subsequent ratings of faculty. In a simulation, 30 managers received hypothetical feedback regarding their own job performance. The managers subsequently evaluated an employee on videotape. Managers who received positive feedback about their performance subsequently rated the employee significantly higher than managers who received negative feedback regarding their own performance. This occurred despite the fact that the managers knew the evaluation of them was bogus. The results of two follow-up field studies involving 74 manager–employee dyads in a manufacturing company in Canada and 39 manager–subordinate dyads in a retail organization in Turkey are consistent with the view that one’s own performance appraisal is related to the subsequent appraisal of one’s subordinates. Both anchoring with insufficient adjustment and a mood induction may explain this effect, but the results are more consistent with the former explanation than the latter.

1. Introduction

The notion that people at work should receive performance appraisals is eminently sensible. This procedure is potentially of enormous benefit to both the individual and the organization, in that the feedback provides people a basis to set goals to both evaluate and improve their performance (Latham, 2007). Why then are both the provider and recipient of this service frequently dissatisfied with the outcome (Greenberg, 1986, 2000)? A major source of dissatisfaction appears to be the difficulty people have in providing an appraisal that accurately reflects a person’s performance over the relevant timeframe (Latham & Mann, 2006). Many sources of this inaccuracy have been documented. They include halo error (Bingham, 1939), similar to me (Wexley & Nemeroff, 1974), contrast and assimilation (Murphy, Balzer, Lockhart, & Eisenman, 1985), and performance-cue bias (Staw, 1975).

The purpose of this study was to explore an additional source of influence in the performance appraisal process. An important, albeit objectively irrelevant, determinant of a judgment about the performance of others appears to be the evaluation received about...
one’s own performance. In short, we hypothesized that when people make formal judgments about the performance of others, those judgments are influenced by the formal performance judgments that others have previously made about them. This is because prior judgments may provide the context and establish the foundation for subsequent related judgments.

There is considerable evidence from behavioral decision-making research indicating that irrelevant information can have potentially large effects on estimates of value under uncertainty. For example, according to the anchoring and adjustment heuristic, people look for a guidepost or anchor when making such estimates (Tversky & Kahneman, 1974). People then make small and often insufficient adjustments to the anchor. Consequently, final estimates are usually quite close to, and strongly biased by, the anchor that was used. This process has been observed in many different situations with a variety of estimates where the value of the quantity being appraised is vague or ambiguous (e.g., Tversky & Kahneman, 1974; Northcraft & Neale, 1987; Whyte & Sebenius, 1997).

Performance appraisals typically involve estimates of value under uncertainty, because the focus of such judgments is the inherently ambiguous notion of an employee’s performance. Among the many sources of uncertainty in this evaluation process is an appraiser’s difficulty recalling performance data at the time of an evaluation (DeNisi & Peters, 1996). Moreover, managers frequently lack the opportunity to observe all relevant dimensions of a person’s performance (Komaki & Citera, 1990). Because of the uncertainty inherent in the rating of performance, managers may be inclined to look for anchors to which small adjustments will be made. The question then arises as to which of the many potential anchors in this situation are most salient to the rater. As noted earlier, we hypothesized that the dominant anchor for the individual doing the appraisal is likely to be the appraisal rating that this person previously received.

In addition to the effect of anchoring, another reason to hypothesize that one’s own previous performance evaluation affects one’s subsequent evaluation of others’ performance is based on research regarding mood. Mood is defined as cycles in one’s feelings (Brief & Weiss, 2002) or transient affective states (Weiss, 2002). One’s own performance appraisal may influence the act of evaluating others by influencing the mood of the rater as he or she rates others. Presumably, a positive rating would translate into a positive mood if such a rating were obtained before, and became salient during, the rating of others. A negative rating might have an even more powerful effect on mood, albeit in the opposite direction. Evidence suggests that raters are too generous when in a good mood and not generous enough when in a bad one (Sinclair, 1988).

1.1. Sources of bias

When rating multiple targets, appraisers often compare the performance of the individual rated second to the performance of the individual rated first. This tendency has been called contrast and assimilation (Murphy et al., 1985). Assimilation occurs when judgments are biased toward the prior impression while contrast occurs when judgments are biased away from the prior impression.

Researchers have investigated the effect of a person’s previous performance appraisal on the evaluation of this person’s present performance (Murphy et al., 1985). Consistent with the notion of anchoring, research on performance cues indicates that raters develop expectations about an individual’s performance. Raters then have difficulty incorporating events that deviate from the employee’s previously observed patterns of behavior.

Both contrast and assimilation and performance-cue biases are relevant to the current study because they indicate that raters often consider irrelevant data (e.g., previous performance ratings or others’ ratings) when making performance judgments. The effect of one’s own performance evaluation on the subsequent appraisal of others is yet to be explored.

1.2. Anchoring and adjustment

In making an estimate of value under uncertainty, people often start at a potentially random anchor and then make only incremental adjustments to it before arriving at their final estimate. As noted above, anchoring and adjustment has been demonstrated in a variety of contexts, including compensation (Chapman & Bornstein, 1996), negotiation (Whyte & Sebenius, 1997), pricing (Mussweiler, Strack, & Pfeiffer, 2000), goal setting (Hinsz, Kalnbach, & Lorentz, 1997), and judgments of self-efficacy (Cervone & Peake, 1986).

In a seminal study, Tversky and Kahneman (1974) asked people to estimate the percentage of African countries in the United Nations after witnessing the spin of a roulette wheel. For one group of participants, the wheel landed on the number 10. For the other group the wheel landed on 65. Each group was then asked whether the actual percentage is higher or lower than the number they saw, and were asked for their own estimate. The median estimate was 25 countries for those who saw the wheel land on 10, and 45 countries for those who saw the wheel land on 65. This occurred despite the fact that people knew the spin of a roulette wheel has no bearing on the correct answer. Tversky and Kahneman attributed this result to a phenomenon they called anchoring and insufficient adjustment.
As in the experiment conducted by Tversky and Kahneman, the anchor used by the decision maker is often not relevant to the required estimate and may even be an obviously random number. An anchor may also be provided by another individual. For example, Chapman and Bornstein (1996) found that an initial request for compensation in a mock trial served as an anchor for the judge’s decision. The more the plaintiff requested, the more compensation the plaintiff received. Similarly, Northcraft and Neale (1987) demonstrated anchoring among real estate agents estimating the price of a home. In this study, the agents received detailed information about the home, including its listing price, which varied from high to low. Despite all the information provided, and despite the agents’ expertise and training in home evaluation, the listing price acted as an anchor; the higher the listed price, the higher the estimate of the value of the dwelling.

Kataoka, Latham, and Whyte (1997) illustrated the effect of anchoring in a selection interview by asking human resource (HR) managers to evaluate an interviewee on videotape. After the HR managers saw the same tape, they were randomly assigned to one of three conditions. Participants in the control group were asked to evaluate an interviewee, using a five-point scale, on multiple dimensions of performance. Participants in the first experimental condition were asked, before evaluating the interviewee, whether the person is a ‘1.’ Managers in the second condition were asked, before evaluating the interviewee, to consider whether the person is a ‘5.’ HR managers who were given either a 1 or 5 anchor evaluated the interviewee significantly closer to the respective anchor than HR managers in the control condition.

In a laboratory experiment, Switzer and Sniezek (1991) examined the role of anchoring and adjustment on predictions of one’s own future performance. Before making such predictions, participants were randomly assigned one of three anchors: (1) an evaluation of their own performance; (2) the performance of others; or (3) an irrelevant value. In all cases, self-estimates of future performance were biased by the anchor that was assigned. Both this study and the study by Kataoka et al. (1997) demonstrate the effect of anchoring and adjustment on the performance evaluation of oneself and an interviewee. In each case, however, the research participants were explicitly provided with an anchor.

Given the many situations in which irrelevant anchors have been shown to bias decision making, it would be surprising if anchoring did not also influence the performance appraisal process. Yet no one to the authors’ knowledge has investigated this phenomenon.

Ability judgments about oneself are inherently egocentric (Weinstein & Lachendro, 1982; Kruger, 1999). When asked to make comparative ability judgments about oneself relative to one’s colleagues, individuals focus on their own abilities rather than considering the abilities of their peers. Kruger and Dunning (1999) found that when participants were asked to estimate their own ability relative to others on an intellectual task, assessments of their own skills were far better predictors than impressions of the skills of their peers. Similarly, Kruger (1999) found that when people compare themselves to others, they focus egocentrically on their own skills and insufficiently take into account the skills of the comparison group. We therefore hypothesized that the results of one’s own previous performance appraisal looms large in the subsequent appraisal of others. That is, one’s past appraisal may be a salient, compelling, and dominant anchor when engaging in the appraisal of another individual’s performance.

To test this hypothesis, we conducted four separate studies using three different methodologies, a case study, a laboratory experiment, and two field studies. McGrath (1982) noted that ‘all research strategies and methods are seriously flawed’ (p. 70). He further argued that rather than making a judgment about which methodology is ‘best,’ research should combine different research strategies and methods into a single research program using ‘multiple means that do not share the same weaknesses’ (p. 80). The aim in research is to maximize generalizability of research findings, precision and control in measures of variables and behaviors, and realism for the participants.

Hypothesis: A manager who receives a favorable performance appraisal subsequently evaluates the performance of another significantly more positively than a manager who receives an unfavorable performance appraisal.

2. Case study

The present authors’ interest in this phenomenon was peaked by the transition of two area coordinators in a business school. The university requires that the Dean and the area coordinator formally evaluate members of the faculty independently on an annual basis. The area coordinator position is typically rotated among the full professors. The faculty are evaluated on their teaching performance (40%), research (40%), and service to the university, scholarly societies, and community (20%). The appraisal scale requires that a composite number ranging from 1 to 7 be given to each faculty member indicative of this individual’s overall performance. Before becoming Dean, the faculty member who held this position had typically received appraisals of 5 over a 20-year academic career. The outgoing area coordinator, a full professor, had received appraisals of 6 or above for the 6 years he had been at the university; the incoming
coordinator, also a full professor, typically received overall ratings of 3 or 4 during that same time period.

Ambiguity existed as to when the incoming professor was to assume the role of area coordinator. Unbeknownst to one another, both the incoming and outgoing area coordinator evaluated the performance of their colleagues, and submitted their respective ratings to the Dean. The consistency in the ranking of the faculty by the three raters, namely, the outgoing area coordinator, the Dean, and the incoming area coordinator was extremely high. For example, where the outgoing coordinator rated a faculty member a 6, the Dean evaluated the faculty member as a 5, and the incoming coordinator gave the person a 4. There was no exception to this trend in the ranking of the 10 faculty members. It appeared as if the appraisal the three appraisers received from their ‘boss’ a year earlier served as the anchor for the appraisal decisions they made regarding their colleagues the subsequent year. This result occurred despite the fact that an area coordinator must defend the performance ratings to the Dean and subsequently, a Dean must be prepared to explain the performance ratings to a Provost. This activity is required for promotion and salary-related issues. Yet, this phenomenon occurred in the presence of relatively objective criteria, namely, number of publications and citations, student ratings of a faculty member’s teaching effectiveness, and number of committees on which a faculty member served.

Case study methods maximize ecological validity, but are limited in that they lack precision, and no conclusions regarding causality can be drawn (Campbell & Stanley, 1963). Hence this finding does not provide a strong basis for concluding that one’s prior performance rating is an anchor for the subsequent rating of others. However, discussions of this finding with the Dean and the two area coordinators led to a formal test of the study’s hypothesis through the use of an experimental design. Therefore, a laboratory simulation was conducted to test the hypothesis that the evaluation a manager receives influences in a systematic way the subsequent appraisal of the performance of another individual.

3. Laboratory simulation

3.1. Participants

The participants were 30 senior level managers in the private sector. Their mean age was 42.11 (SD = 6.94); 21 were male.

3.2. Procedure

Each manager received hypothetical feedback concerning his or her performance during the past year, as well as a job description of a direct report. The managers knew that the evaluation of them was fictitious. Managers read the job description after having been asked to assume that the direct report had been performing his job for the past year.

The managers were randomly assigned to one of two performance feedback conditions. Managers in the positive feedback condition were told the following: ‘You have exceeded your boss’s expectations and everyone is extremely pleased with your progress. You have outstanding management skills that are only surpassed by your interpersonal skills. Your peers find your performance to be highly effective. They look to you for direction, and they respect your opinion greatly. Your subordinates think of you as a mentor, and look to you for direction on matters beyond the scope of your portfolio. In summary, you are performing well above average. You have a strong and secure future with the company. Your overall rating is 7 on a seven-point scale.’

Managers in the negative feedback condition read the following: ‘Your boss is disappointed with your progress and needs to see significant improvement in your performance in the near future. There is widespread agreement that you have poor management skills. Moreover, your interpersonal skills are lacking. Your peers find your abilities to be medioc approached. They complain that they are constantly ‘cleaning up your mess.’ Your subordinates do not respect you and find it difficult to take your direction seriously. They always double check your instructions with your boss. In summary, you need to improve significantly in order to ensure a future with the company. Your overall rating is 1 on a seven-point scale.’ The managers were unaware of any difference between the feedback and rating of performance in the two conditions.

Subsequent to receiving feedback, the Positive and Negative Affect Schedule (PANAS) mood assessment scale was administered (Watson, Clark, & Tellegen, 1988). The scale consists of 10 positive affects (interested, excited, strong, enthusiastic, proud, alert, inspired, determined, attentive, and active) and 10 negative affect inents (distressed, upset, guilty, scared, hostile, irritable, ashamed, nervous, jittery, and afraid). Participants were asked to rate the extent to which they felt these affects on a scale from 1 to 5, based on the strength of emotion where 1 = ‘very slightly or not at all,’ and 5 = ‘extremely.’ Watson, Clark, and Tellegen (1988) demonstrated that the scale has adequate internal consistency (ranging from .84 to .90) and test–retest reliability. This test yields a score for both positive affect and a score for negative affect.

Managers next assessed the performance of the individual whose job description had been provided to them. Consistent with Mero and Motowidlo (1995), the employee’s performance was shown on videotape to hold performance constant across conditions. There is
no significant difference in rating accuracy of a live vs a videotaped performance (Lifson, 1953; Ryan, Daum, Bauman, Grisez, & Mattimore, 1995).

The following instructions were provided to the participants in both conditions: 'Review the job description and then evaluate this person’s performance.'

A white male was the only individual visible on the videotape. He discussed his performance for 4 min with an individual who was off camera. Consistent with the evaluation form used in the case study, the managers were then asked to evaluate the job performance of this individual on a seven-point (1 = low, 7 = high) scale. On this scale, 1 was positioned on the left side and 7 was positioned on the right. Each number from 1 to 7 was spaced at regular intervals along a straight line.

3.3. Results

To test the effectiveness of the manipulation of the independent variable, the mood of the managers, as previously noted, was assessed immediately following the feedback they received, and immediately before them assessing the person shown on the videotape.

Managers who received a negative appraisal (x = 30.92, SD = 6.65) reported significantly higher negative affectivity than managers who received positive feedback (x = 15.77, SD = 8.44; F = 43.38, p < .001). Moreover, managers who were given positive feedback (x = 40.14, SD = 6.98) reported significantly higher positive affectivity scores than those who received a negative appraisal (x = 23.47, SD = 6.65; F = 18.63, p < .001). Thus the manipulation appears to have been effective.

The mean rating of the individual was 2.13 (SD = 1.46) and 3.14 (SD = 1.30) in the negative and positive appraisal conditions, respectively. Analysis of variance supported the hypothesis that a manager’s own performance rating influences the performance rating he or she subsequently gives to others (F = 3.60, p < .05). These results are consistent with the notion of anchoring and adjustment.

A rival hypothesis is that a mood induction from the raters’ own performance appraisal explains the results. Consequently, ANOVA was conducted. Neither positive (F = 6.88, p > .05) nor negative affectivity (F = .58, p > .05) significantly predicted the appraisal ratings. Therefore, mood was ruled out as a significant predictor of performance appraisal ratings in this study.

3.4. Discussion

A laboratory experiment maximizes control. Internal validity is high (Campbell & Stanley, 1963). Nevertheless, it may have this advantage at the expense of limiting realism for the participants. Hence, generalizability of the findings to an organizational setting may be low. The managers in the present simulation were not accountable to anyone for their performance evaluation of the videotaped employee. Consequently, two field studies were conducted, in part to further minimize the chance that a mood induction could be responsible for the effect observed in the simulation, and in addition to determine whether the findings from the case study in a university, and those from the simulation generalize to organizational settings. Specifically, the results of actual performance appraisals conducted by managers who had received their own appraisal a year earlier were examined.

4. Field study 1

4.1. Manufacturing firm

4.1.1. Participants

The participants were 27 managers employed by a manufacturing company for at least 5 years. Their mean age was 41.30 (SD = 2.51) years; 20 were male. All of the managers were very familiar with the company’s evaluation system. This system required managers to evaluate an employee’s performance across multiple dimensions of performance using a three-point scale, and to formulate a composite score of each employee’s performance for administrative and developmental purposes.

4.2. Procedure

As part of the organization’s performance appraisal system, the managers received formal evaluations of their performance from their superiors at time 1. These data were collected and archived by the researchers. The researchers were not involved in the organization’s performance appraisal process in any way. At time 2, 1 year later, the 27 participant managers evaluated their 74 employees. Again, the researchers were not involved in this process in any way. They merely received the performance appraisal data from the organization’s HR manager after the appraisals had been conducted. No one in the organization was aware of the hypothesis that was being tested.

4.3. Results and discussion

The performance evaluations received by the managers predicted their performance evaluations of their employees 1 year later. The Spearman ρ was .25 (p < .05). The evaluation received by the manager who did the appraisal was a significant predictor of the manager’s evaluation of an employee a year later.
This study was conducted in Canada as were the case and laboratory study. The results of this correlational study were replicated in a Muslim country to further assess the generalizability of the findings supporting our hypothesis.

5. Field study 2

5.1. Retail organization

5.1.1. Participants
The participants were managers and subordinates employed by a retail company in Istanbul, Turkey. All of them had received training by the company on how to use the company’s evaluation system. Data were available on 39 managers and 227 manager–subordinate dyads. One employee was selected for each manager using a random number generator.

5.2. Procedure
The managers received formal evaluations from their superiors. These data were collected and archived by the researchers. As was the case in the first field study, the researchers were not involved in the organization’s performance appraisal process in any way. The managers subsequently evaluated their employees. We received the performance appraisal data from the organization’s HR manager after the appraisals had been conducted. As in the previous field study, no one in the organization was aware of the hypothesis that was being tested.

5.3. Measure
The organization’s performance appraisal instrument was used in this study. Again, the authors were not involved in any way in the selection or development of this instrument. Employees are measured on two dimensions in this organization. The first dimension was a set of competencies relevant to a retail sales environment. In total, six competencies comprise the first dimension, namely, honesty, customer orientation, team playing ability, communication skills, expert knowledge about the products, and personal motivation. Each competency was assessed using between two and 15 items. For example, an item assessing the competency customer orientation was ‘Welcomes and sends off customers in a kind and friendly manner.’ A sample item for measuring team playing ability was ‘Shares knowledge with peers and supervisors.’ An item under communication skills was ‘Uses clear and specific language.’ Each item was evaluated on a five-point Likert-type scale where 0 indicated ‘very poor’ and a 4 designated ‘excellence.’ The scores for all competencies are averaged by the company to develop a mean competency score.

The second dimension assessed personal objectives. This is because the organization has an established goal-setting program. Employees are evaluated on the extent to which they have attained their goals using a five-point Likert-type scale from 0 ‘failed’ to 4 ‘extraordinary.’ The scores for each objective are summed to develop a total objectives score.

An employee’s overall performance evaluation in this company is the sum of the total competency score and the total objectives score. Overall evaluation scores in this organization can range from 1 to 5. Consequently, the composite score for each individual was used in the present study as this is the score used by the company for making administrative decisions.

5.4. Results
The mean rating given to a manager was 2.74 (SD = .60). The mean rating given to an employee by a manager was 2.85 (SD = .66).

The time period between a manager’s appraisal and the subsequent appraisal of a subordinate ranged from 4 to 6 months. The two data sets were significantly correlated (r = .57; p < .01). The performance rating received by managers predicted their subsequent rating of employees.

6. Discussion
This research relied on three different methodologies, namely a case study, an experimental simulation and two correlational studies in field settings. This research involved participants from three different organizational settings, namely, area coordinators in an academic institution, managers in a laboratory, managers in a manufacturing plant in Canada as well as managers in a retail organization in Turkey. Three of these studies occurred in the West, namely Canada; one occurred in the East, Turkey. Studies 1, 2, 3, and 4 also used different appraisal instruments. As McGrath (1982) noted, ‘using multiple methodological probes to gain substantive convergence by methods . . . compensate for one another’s vulnerabilities’ (p. 80). For all of the limitations inherent in each of the four studies, the results across all four are nevertheless consistent. They provide support for the hypothesis that one’s own performance appraisal affects one’s subsequent appraisals of others. The appraisal a manager receives affects or predicts the subsequent performance appraisal that a manager gives to others.
The present results suggest a modification to a psychological axiom, namely that the best predictor of future performance is past performance. The present study suggests that a predictor of one's future performance evaluation is the prior performance appraisal of the individual who will be doing the assessment. The results of the two correlational studies in the field indicate that this effect occurs even when the manager's and the subordinate's rating are separated by a considerable period of time. These results are consistent with earlier findings indicating inaccuracy in performance appraisals due to rater bias (e.g., Lance, 1994), and suggest yet another reason why managers should be careful when relying on the outcomes of this process to make HR related decisions (e.g., promotion, demotion, compensation). The practical significance of these data for an employee is that they suggest the benefit of seeking a boss who is viewed as a high performer so that one's performance is likely to be evaluated more highly than is the case where one's boss is a seen as less than stellar within the organization.

A complementary hypothesis to anchoring as an explanation of this phenomenon is that a negative mood is typically induced upon receipt of a poor performance evaluation, and a positive mood is induced upon receipt of a good one. In general, a negative mood on the part of the rater is correlated with negative evaluations of others, and a positive rater mood is correlated with positive evaluations (Sinclair, 1988; Longenecker, Jaccoud, Sims, and Gioia, 1992).

In the present laboratory simulation, a mood change did not account for the differential evaluations of the videotaped employee. Moreover, in the case study as well as the two field studies, the evaluations received by the area coordinators and managers and the subsequent evaluations given to their subordinates were separated in time. This temporal gap would appear to eliminate the rival hypothesis that a mood induction explains the rating patterns observed across the four studies. Moods are fleeting, short-term occurrences (Brief & Weiss, 2002).

Another alternative to the notion of anchoring on one's own performance rating when evaluating others is the possibility that the managers who were rated highly in the two field studies received this rating because their subordinates were indeed high performers. Similarly, the managers who received low evaluations from their superior and subsequently gave lower evaluations to others a year later may have done so because they had subordinates who, in fact, performed poorly. Thus, the results of the two field studies may reflect reality. The laboratory experiment, however, controlled for this possibility by having all raters evaluate the same individual at the same point in time. Moreover, this rival hypothesis does not explain the finding reported in the case study.

The contributions of this research to the literature are at least twofold. First, the results demonstrate the existence of a previously unidentified and potentially pervasive effect in the performance appraisal process. This discovery is significant because performance appraisals provide raw material for decisions that are important to both the individual and the organization. Second, this research suggests that this phenomenon is due to the effect of anchoring and adjustment on performance appraisals.

Although the phenomenon of anchoring in judgment and decision making has been demonstrated in many contexts, there is currently no substantive theory of anchor selection. The capacity to predict which anchor people will use is important, however, because this knowledge is necessary to make accurate predictions about people's estimates of value under uncertainty. Whyte and Sebenius (1997) had suggested that when multiple anchors of varying quality are present, individuals are inclined to select the anchor that is most relevant to the task at hand. The present findings are consistent with their view that the most compelling anchor to a decision maker is not necessarily the most logically relevant, but rather the most personally significant. In this instance, the appraisal one receives from one's boss.

Because participants in the current research were seasoned managers, the results indicate that susceptibility to anchoring is not moderated by a person's experience. Nor does susceptibility to anchoring appear to be mitigated by the amount of information that managers have about the performance of their subordinates. Some research (e.g., Mussweiler & Strack, 2000) has demonstrated that the more a judge knows about something, the less inclined this person will be to rely on a heuristic when evaluating it. This behavior, however, was not apparent in either the case study or in the two correlational studies where managers were required to make judgments about their subordinates' performance, something about which they had considerable information over an extended time period.

A major limitation of these findings is lack of theory to explain them. Future research may benefit from examining this phenomenon within the context of organizational justice or social exchange theory. Anchoring based on the evaluation one receives may serve as benchmarks for what one should consider as fair, and/or what one's supervisor 'expects.' This perception quite conceivably could be taken into account when performing subsequent evaluations of one's subordinates.

A second limitation of this research is the lack of 'true scores.' The use of true scores, and the concomitant use of difference scores, can be problematic (Edwards, 2001). Nevertheless, lack of knowledge of true performance makes it arguable whether the an-
The anchoring effect is error, irrelevant, or bias. Logical inferences, however, can be drawn. Ratings affect a faculty member’s salary increase in the university described in the case study. A faculty member cannot be simultaneously a 6, 5, or 4. One, or more, of the three raters must be wrong. A person viewed in a videotape by the same managers at the same point in time cannot be both a high and a low performer. One or both groups of managers must be wrong. The two correlational studies do not allow inferences regarding causality. They only provide evidence that the phenomenon occurs, namely the evaluation one receives predicts the evaluation one subsequently gives others.

Taken together, the results from these four studies suggest that future research is now required on ways to overcome this phenomenon. Training programs (e.g., Heslin, Latham, & VandeWalle, 2005), in addition to the use of a highly structured behavioral appraisal instrument, may be effective in reducing this effect (Fay & Latham, 1982). Similarly, keeping an ongoing diary of behavioral observations of an employee may prove to be effective in minimizing it (DeNisi & Peters, 1996).

Future research should investigate the extent to which the present findings relate to the performance evaluation of a CEO. If a manager’s performance appraisal affects the manager’s judgment of an employee’s performance, it is conceivable that yet another source of influence on a performance appraisal is judgments of an organization’s performance. A firm’s overall performance may act as an anchor in the performance appraisal process. An organization’s performance may vary even when an individual’s performance remains relatively stable. For example, manufacturing firms in Canada are currently being hurt by the rise of the Canadian dollar to the American. Exports from Canada to the United States have dropped. Whether a firm’s performance, due to economic conditions or fluctuations in monetary currency, inappropriately influences the performance appraisal of a CEO is a worthy topic for future research.

References


