

# COGNITIVE APPROACHES TO PERFORMANCE APPRAISAL

Angelo S. DeNisi and Kevin J. Williams

---

## INTRODUCTION

The question of how to accurately measure performance is by no means novel in fact, it forms the cornerstone for much of human resources management. Yet while organizational scientists have been trying to construct accurate appraisal devices since the turn of the century, bias in ratings continues to exist. In what could be seen as a reflection of approximately 70 years of research, the introduction of the graphic rating scale coincided with Thorndike's (1920) demonstration of "a constant error in psychological ratings," that is, halo bias. Since then there has been an alternation between the development of new evaluation systems and documentation of rating bias. Although the problem has never changed, researchers have changed how they go about addressing it. The present paper reports on the most recent attempts to deal with the pervasive accuracy problem in performance appraisal.

---

Research in Personnel and Human Resources Management, Volume 6, pages 109-155.  
Copyright © 1988 by JAI Press Inc.  
All rights of reproduction in any form reserved.  
ISBN: 0-89232-856-8

At the same time that organizational scientists have been addressing the issue of rating accuracy, formal performance appraisal systems have become an integral part of organizational life. By the middle of the century, approximately half of the companies surveyed by the National Industrial Conference Board reported having employee performance appraisal systems (Tiffin & McCormick, 1965); current estimates indicate that the percentage continues to increase, exceeding 70% for certain jobs (Dunham, 1983). Considerable time and effort go into appraisal preparation, the actual appraisals themselves, and the subsequent feedback and decision making based on the results of appraisals. Important decisions concerning pay increases, promotions, and terminations are often based on judgments concerning how well someone is performing a job. Yet, despite the important role appraisals now play in organizations, it is clear that we still do not know much about the appraisal process. Despite the research that has been done on appraisals, little has been done to improve the accuracy of ratings, and rating problems have shown themselves to be extremely resistant to efforts to eliminate them. Further, if one consults the recent review by Bernardin and Villanova (1986), one might actually claim these problems persist in spite of that research.

Why do these problems persist? Perhaps the answer lies in the approaches that have generally been adopted to deal with them. It is not our purpose to provide an exhaustive review of the appraisal research conducted over the past 60 years or so, but if one were to examine that literature (e.g., Bernardin & Beatty, 1984; DeNisi, Cafferty, & Meglino, 1984; Feldman, 1981; Landy & Farr, 1980), it would become clear that the majority of this research has focused on either the development of "better" rating scale formats, or the training of raters to help them avoid rating errors. Unfortunately, the failure of these approaches to lead to any real improvement in appraisals has been duly noted. Frank Landy and Sheldon Zedeck, in their introduction to the conference, "Performance Measurement: Directions for the Future" (Landy, Zedeck, & Cleveland, 1983), described vividly the "quagmire of methodology" that performance appraisal research wallowed in from 1930 to 1970. Very little was learned during this time about *how* raters reached their judgment. The failure of these approaches led Landy and Farr (1980) to suggest that researchers study rater cognitive processes associated with appraisal decisions as a means of generating "more substantive propositions concerning where rater biases come from, why certain processes minimize rating errors whereas other procedures exaggerate these errors, and so forth" (p. 96). As we will see, some research activity during the 1950s did actually address these cognitive issues (Wherry, 1952). However, this research was clearly in the minority, and largely overlooked, perhaps because the cognitive "revolution" in psychology had not yet taken place.

The present paper is concerned with the models and research that have followed from Landy and Farr's (1980) call. First, we will attempt to outline the major cognitive models of the appraisal process and to discuss studies based upon these models. We will then turn our attention to a program of research based upon one

of these models to illustrate how such research might progress. We will conclude by reviewing the sources of bias that have been identified and what processes or interventions might be pursued. In our view, cognitive approaches to studying performance appraisal have already yielded some practical suggestions for ways to improve the process. Given continued programmatic research, however, the potential yield of these approaches is even greater.

## COGNITIVE MODELS OF THE APPRAISAL PROCESS

It is difficult to fix a date when cognitive approaches to the appraisal process began. For example, Mosier (1940) conceptualized evaluations in terms of psychophysical scaling of stimuli, and was able to demonstrate that many of the ideas from mental testing, such as the notion of reliability, applied to ratings. These ideas were later incorporated into Wherry's (1952) model of the rating process, where many sources of rating unreliability were seen as evolving from cognitive processes. Wherry (1952) also drew upon Bartlett's (1932) work on human memory in his model. Specifically, Bartlett (1932) suggested that we tend to remember events according to a generalized pattern or "schema," and that reliance upon these schemata lead to our forgetting details inconsistent with the pattern and falsely "recalling" details consistent with it. This notion has become an important part of more recent models as well.

### Wherry's Model

Wherry (1952) presented the first formal cognitive model of performance appraisal. Although his original model was never published, much of our recent work was truly anticipated by Wherry. A tribute to the durability of his contribution is the fact that 30 years later, an edited version of his original technical reports was seen as a sufficient contribution to warrant publication (Wherry & Bartlett, 1982).

Wherry's (1952) model relied heavily upon test theory and began by suggesting that a performance incident exhibited by a ratee was due to the ratee's ability and some random error component, as well as any number of environmental factors. He then expanded his model to consider specific and general rater biases, schemata, and other biasing effects on the rater's recall of ratee performance, and similar effects operating to bias the rater's initial perception of the ratee's performance. Much of the remainder of Wherry's work suggested factors which would influence the relative importance of different sources of bias in ratings. Although we know of no studies that were designed specifically to test the theorems and corollaries concerning these sources of bias, many of these theorems have re-emerged later to receive attention under the auspices of "newer" models.

For example, Kane and Lawler's (1979) suggestions concerning Behavioral

Discrimination Scales addressed the potentially confounding effects of frequency of occurrence of behaviors on the evaluation of those behaviors. This problem was also noted earlier by Wherry (1952), who's Corollary 5a states: "Rating items which refer to frequently performed acts will be rated more accurately than those which refer to acts performed rarely or at long intervals" (Wherry & Bartlett, 1982, p. 533). Theorems 6 and 7 also anticipated more recent views on rater training which de-emphasize training to avoid errors (Bernardin, 1979; Pulakos, 1984, 1986; Thornton & Zorich, 1980) by suggesting also that raters be trained in what to look for and how to observe behavior more objectively. Other theorems or corollaries predate recent work, suggesting that raters be provided with the scales they are to use before observation of behavior (DeNisi & Summers, 1986); that raters may be reluctant to give poor but accurate ratings when they will have to justify those ratings to ratees (Fisher, 1979; Sharon & Bartlett, 1969); and that observation with the purpose of recall will facilitate recall (although recent work in the social cognition area indicates that such memory instruction sets may result in less recall than impression formation sets, e.g., Hamilton, Katz, & Leirer, 1980). Thus, as we shall continue to see, Wherry's (1952) cognitive model set the stage for much of what was to follow almost 30 years later.

#### Feldman's Model

As was noted earlier, Wherry's (1952) model was largely overlooked in the literature for many years. In fact, the next major cognitive model does not even reference Wherry's work (Feldman, 1981). Feldman (1981) drew upon a growing body of literature in social cognition to develop a model that is, in essence, an applied person perception model. His model begins by recognizing that there are many demands on a rater's time in addition to appraisal-related tasks and so characterizes the appraisal context as an "uncertain, informationally 'noisy' environment" (p. 128). Given this, Feldman noted that a rater must engage in the following cognitive processes before completing an appraisal: (1) recognize and attend to relevant information, (2) organize this information in memory for later access, (3) recall this information in an organized manner when ratings are required, and (4) integrate this information into a summary judgment. His model moved beyond simpler attributional models which considered only information that was "complete and immediately available" (p. 128).

Perhaps the major feature of Feldman's (1981) model is the consideration of automatic and controlled processing of information (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Under conditions of automatic processing, a rater encounters a stimulus with which he or she is quite familiar or has had considerable experience. The processing of this stimulus proceeds without conscious awareness on the part of the rater, and does not interfere with ongoing mental activities (Posner & Snyder, 1975). According to this view, upon observing a

ratee performing a given task, the rater would "automatically" process behaviors as either relevant or irrelevant to the appraisal decision and may even categorize the performance as good or poor, without consciously attending to the encoding process. Automatic processing would also be in operation whenever a rater encountered a piece of information about the ratee that was consistent with the rater's expectations about that ratee. Controlled processing of information, on the other hand, is under the conscious control of the individual and is seen as intentional. Controlled processes may operate, for example, when a rater encounters information that is inconsistent with expectations or when its relevancy to the appraisal task is uncertain. This requires active attention on the rater's part and so necessitates controlled processing. We might consider the situation of driving home from work as an analogy for this view of automatic and controlled processes (we will discuss alternative views later). We take the right exit and make the correct turns without even being fully conscious of it. One day, however, when we go to take the usual exit we find it closed because of construction and so we must take an alternate route. Taking this alternate route requires our attention and concentration because it necessitates a deviation from our normal routine.

Feldman (1981) proposed that reliance upon automatic processes results in categorization (cf. Rosch, 1975) of performance incidents and ratees. That is, incidents and eventually ratees are assigned to a category, such as good, poor, lazy, ambitious, male or female, and future processing of information is based on the characteristics of that category. Thus, other information about that person is interpreted in a manner consistent with the category to which he or she has been assigned. The assigned category will also serve to screen out certain pieces of information as irrelevant and, in essence, the appraisal made is a function of a stereotype the rater holds about a certain category of people. It is only when controlled processing is evoked (i.e., when an observed incident differs significantly from the performance expectations for that category) that the rater consciously attempts to make some attribution for that incident which may or may not lead to the assignment of that person to a new category. In either case, however, the person is categorized and the category chosen determines much of what else occurs in the appraisal.

Feldman's (1981) model also presents suggestions for future research and applications. Generally, he has suggested we move away from research on rating scales, except that certain types of rating scales might make some categories more accessible to raters than others (recall the similar proposal from Wherry). Although this issue has not received much empirical attention, Feldman (1986) has developed a more complete framework for studying rating scales from a cognitive perspective. It is quite likely that this more fully articulated model will lead to a stream of empirical work. Feldman (1981) also discussed the need for research on other cues for category salience and the role of categorization on other cognitive processes. A study conducted by Feldman and his associates (Feldman, Camburn, & Gatti, 1986) generally failed to find support for several

hypotheses related to categorization but, as we shall see later, there has been extensive research in other areas that owes some of its theoretical justification to Feldman's (1981) model. Specifically, Feldman's discussion of automatic and controlled process in evaluations may be one of the more important contributions of his work.

### Ilgen and Feldman's Model

Feldman's (1981) model was refined and more fully specified a few years later by Ilgen and Feldman (1983). This later model emphasized the work context more than the earlier work did. Specifically, it considered the function which the appraisal is intended to serve, opportunities for observation, problems associated with continuous work groups (i.e., where it is difficult or impossible to rate the performance of an individual independent of the group's performance), and limited opportunities for rewarding performance. The authors, consistent with Feldman's (1981) work, emphasized the role of cognitive categorization, but then went on to specify different processes for automatic and controlled categorization. It is interesting to note that in their controlled categorization model, they included an active search for information which, as we shall see, is a major factor in the DeNisi, Cafferty, and Meglino (1984) model.

Ilgen and Feldman (1983) also have focused on the rater much more than have other cognitive theorists, and they have drawn upon the cognitive literature to reinforce some suggestions for practice that had been made earlier without the benefit of research on cognitive processes (e.g., the separation of feedback and rewards suggested by Meyer, Kay, & French, 1965). They closed their discussion, though, with a series of very cognitively oriented recommendations for future research and practice dealing with training and appraisal instrument development. Although many of these specific recommendations are worthwhile, we think that a great deal of what they recommended can be distilled into a paragraph Ilgen and Feldman (1983) used early in their discussion of recommendations:

Let us first state three well accepted assumptions, then consider the implications of these assumptions. First, assume that the behavior of the employee is the most objective individual contributor to the individual's performance. Second, assume that appraisers are able to observe and evaluate behaviors. Finally, assume that appraisers use some category system to store information about another's performance. Given these conditions, accurate appraisal is most likely if we understand (1) the relevant behavior set, (2) the relevant category set, and (3) we either develop rating scales that closely fit the match of these two, or we train people to use a new category system that incorporates (1) and (2) (p.181).

We conclude our discussion of the Ilgen and Feldman (1983) model with this quote because, not only does it outline the philosophy of those authors concerning how cognitive models can be used in practice, but it also represents the view of other cognitive researchers as well. As we shall point out again later, the

differences among these researchers and theorists is not in the general approach taken, but in its specific emphases.

### *A Note on Automatic and Controlled Processes*

Performance appraisal researchers seem to have wholeheartedly embraced the notion of automatic and controlled processes (in addition to Ilgen & Feldman, 1983, see Lord, 1985). This two-process conceptualization of information processing has provided the theoretical base that has been missing from much of the cognitive work in the area. As such, automatic and controlled processes have been useful in explaining a number of rating phenomena. However, before we rely too much on these terms to explain heretofore unrevealed processes, we need to clarify exactly what these terms are describing. Although performance appraisal models have mainly drawn from the work of Shiffrin and Schneider, other models exist and have slightly different sets of assumptions (Bargh, 1984; Logan, 1980; Posner & Snyder, 1975). Few attempts have been made to integrate these models as they relate to performance evaluation. Without the theoretical understanding of the two-process model, and given the general appeal of this approach, we may find ourselves applying automatic and controlled processing notions in an inappropriate manner, or to areas in which they are not truly relevant. The two-process model may be analogous to what is occurring in certain rating situations, but may not be the actual explanatory mechanism.

Recently, Bargh (1984) has synthesized the work on automatic and controlled processes, emphasizing what is and is not reflective of these processes and reviewing a number of research applications. Of particular importance to performance appraisal may be his statements of what automaticity is not. Bargh feels that "there is a very real danger that the 'automatic' label has been so widely applied that the distinction may cease to have any real meaning" (p. 3). Specifically, Bargh stated that automatic processing does not necessarily imply (a) irrationality, since decision makers do not make the best use of data even when "consciously" processing it, or (b) lack of awareness, since individuals lack awareness of most of their cognitive processes, whether automatic or controlled. Bargh also offered the following definition for the two-process model:

Automatic processes can operate in parallel, are not affected by how much of the limited capacity is currently available, and are not under the person's control. Conscious or controlled processes, conversely, are serial in nature, are capacity-limited, and are intentional on the part of the individual. Automatic processes develop out of considerable and relatively consistent experience with an environmental object or event, and so are comparatively fixed and inflexible, while control processes are adaptable to novel and unusual situations (pp. 12-13).

It is not always clear that these conditions have been met when the automatic/controlled concept has been applied. Later, we will introduce the concept of "on-line" judgments (Hastie & Park, 1986) and argue that these judgments seem

to share much in common with "automatic" processes in performance appraisal. Perhaps some theoretical integration is possible and could be helpful in this area.

### DeNisi, Cafferty, and Meglino's Model

DeNisi, Cafferty, and Meglino (1984) proposed a model that is, in large part, similar to the Ilgen and Feldman (1983) model and, since we will discuss the DeNisi et al. (1984) model again later, the particulars are not reviewed here in any detail. There is, however, one major difference that needs highlighting. Although DeNisi et al. (1984) considered the same sub-processes as the others, there is a greater emphasis on the information acquisition activities of the rater. According to DeNisi et al., the pattern in which information is acquired by the rater has a significant impact on how that information is later processed. This approach assumes that factors such as the purpose for which an appraisal is conducted, prior impressions about ratees (i.e., expectations), and time constraints will all affect the way in which raters go about acquiring information. Consistent with Feldman's (1981) notion that the rater has many competing demands on his or her time, the rater is forced to make evaluation decisions based only upon samples of ratee performance. The particular sampling strategy chosen by the rater will affect the actual information available, as well as the pattern in which that information is acquired.

Accordingly, much of the research based on the DeNisi et al. (1984) model has focused on purposeful acquisition of information by the rater. The emphasis has been on rater activities in situations where the rater intentionally searches for performance information in order to make a judgment. Studies that investigated these search processes not only found evidence of systematic search strategies, but also demonstrated that these strategies were sensitive to constraints on the amount of information a rater could acquire (which represents time constraints) and the purpose for which the appraisal is conducted (Cafferty, DeNisi, & Williams, 1986; DeNisi, Cafferty, Williams, Blencoe, & Meglino, 1983; Williams, DeNisi, Blencoe, & Cafferty, 1985).

The DeNisi et al. (1984) model also emphasized the importance of patterns of information storage in memory. Working from earlier work on memory models in social cognition (e.g., Wyer & Srull, 1981) these authors suggested that information might be differentially accessible depending upon how it was organized in memory. This notion can be seen as expanding on Feldman's (1981) categorization process. In fact, the level of organization in which DeNisi et al. (1984) are interested may precede or give rise to the use of social categories in memory; they examine how individuals initially organize information in memory, and how they subjectively organize or "compartmentalize" information (cf. Pryor & Ostrom, 1981; Srull & Brand, 1983). It is posited that certain patterns of organization will increase recall and rating accuracy.

Support for this suggestion, and for the role of acquisition strategy in memory

organization, was obtained in several studies (Cafferty et al., 1986; Williams, DeNisi, Meglino, & Cafferty, 1986). As a result, these researchers proposed that a key to understanding the appraisal process is the ability of the rater to organize information in some meaningful pattern in memory. This ability is even more critical when one considers that performance information is often acquired incidentally as the rater goes about performing other tasks. Only if this information can be organized in memory can it be recalled and used effectively in appraisal decisions. Although Wherry (1952) did not specifically address the organization of information in memory, it is again interesting to note that his model pays considerable attention to the critical nature of recall in determining the accuracy of ratings.

### Other Models

We have emphasized these four cognitive models of the appraisal process because of their historical role and/or because of the research they have generated. There have also been a number of other cognitive models, including one in the leadership area, that deserve attention because of their implications for the appraisal decision process.

DeCotiis and Petit (1978) proposed three major determinants of rater accuracy: rater motivation, rater ability, and the availability of judgmental norms. Rater ability is tied closely to the rater's opportunity to observe relevant behavior and is not as sophisticated as the treatment of ability in other models. Also, their concept of standards is concerned more with the congruence between rating scale content and job content than with the availability of actual standards or norms. The major contribution of this model lies in its recognition of the importance of rater motivation. The authors saw motivation as being a function of perceived consequences of ratings, as well as the perceived adequacy of the appraisal instrument. Although some recent work has tried to discuss the cognitive implications of rater motivation (DeNisi & Williams, 1986), cognitive models have generally focused exclusively upon rater ability to the detriment of motivational considerations. Future research should attempt to relate motivational consequences to the processing of performance information (cf. Bernardin & Villanova, 1986).

Borman (1978) proposed a rudimentary cognitive model which attempted to establish how accurate rates could be in the best of situations. He proposed that rating accuracy was a function of: (1) rater opportunity to observe relevant behavior, (2) rater's knowledge of rating errors and their role in accuracy, (3) the quality of the rating instruments used, and (4) organizational constraints on raters which prevent them from recording their best estimates of ratee performance. Actually, this paper is probably less noteworthy for the model itself than for the associated study. Borman (1978) reported that although raters in his study had almost every advantage (except for a longer exposure to ratees and an in-depth training program), there were still problems with their ratings. Borman (1978)

suggested that processes other than those considered in his study could be influencing rating accuracy. Many of the causes of rating inaccuracy about which he speculated were later considered as part of other cognitive models.

Cooper (1981b) presented a model which, although it primarily addressed halo error, provided a sequential view of the rating process. Cooper suggested that rating accuracy depends on the following: (1) observation of actions; (2) observation, encoding, aggregation, and storage in short-term storage; (3) short-term memory decay; (4) transfer to long-term storage and aggregation; (5) long-term memory decay; (6) presentation of categories to be rated; (7) observation and impression retrieval from long-term storage; (8) recognition of observations and impressions relevant to rating categories; (9) comparison of observations and impressions to rater's standards; (10) incorporation of extraneous considerations; and (11) ratings. Cooper (1981b) noted that halo error can be introduced in every step in the process, but it is clear that other sources of inaccuracy can also be introduced at each step. One of the major propositions from this model was the notion that conceptual similarities among rating dimensions are used by raters when making judgments. Conceptual similarity, in turn, was thought to lead to overestimates of the true level of intercorrelations among dimensions, thus resulting in halo. Kozlowski, Kirsch, and Chao (1986) recently found support for this hypothesis, but only when raters were unfamiliar with the job and/or ratees.

Cooper (1981a) argued that halo error could be reduced by increasing rater familiarity with raters. Specifically, familiarity should allow the rater to obtain information to disconfirm perceptions of covariance. Obviously, the results of Kozlowski et al. (1986) support this position. On the other hand, there are researchers who have suggested that familiarity tends to result in a rater organizing information in memory according to person categories. This may be a problem because Cafferty et al. (1986) reported that organizing by persons actually produced less intra-ratee variability in ratings (and thus more halo) than did alternate organizing schemes. Cooper (1981b) further argued that the longer the time between observations and rating, the more reliance there would be upon memory-based judgments and so the more halo would be present. Cooper (1981a) found support for this proposition in a related study. Further support for this premise was reported by Murphy and Balzer (1986), who also found that memory-based ratings were more accurate than immediate ratings, indicating that the formation of global judgments might actually aid accuracy even though halo is increased. One other suggestion by Cooper (1981b) was to have raters rate all ratees on one category before moving on to rate all ratees on the next category and so on. Although Cooper noted that tests of this approach have generally been unsuccessful (Brown, 1968; Blumberg, DeSoto, & Kuehe, 1966; Johnson, 1963; Taylor & Hastman, 1956), the suggestion may bear reexamination in light of recent findings. Cafferty et al. (1986), for example, found the most accurate ratings in conditions where raters organize information in memory around task categories. It has been suggested (DeNisi et al., 1984; DeNisi & Summers,

1986) that introducing rating scales prior to observation helps raters use the rating scale format as a cue for organizing information in memory. This would lead us to hypothesize that an early introduction of the type of scale mentioned by Cooper might increase accuracy. In the early studies cited, scale format was introduced only at the time ratings were required, which may not affect organization in memory. Thus, Cooper's (1981b) proposition that "task-blocked" ratings decrease halo awaits a more appropriate test.

Before closing this discussion, it is necessary to note a model that was actually proposed to explain how subordinates evaluate leader behavior, but which can be applied to the measurement of performance. Robert Lord and his associates (Lord, 1985; Lord, Foti, & DeVader, 1984; Lord, Foti, & Phillips, 1982) have argued that people have fairly consistent prototypes about the traits and behaviors that characterize effective leaders. Perceptions of potential leaders and evaluations of actual leaders are seen as being a function of the observer's perceptions that the target person exhibits these prototypical behaviors or possesses these traits. This line of reasoning is not dissimilar to that proposed by Feldman (1981) and could easily be applied to performance appraisals where raters would have prototypes of effective and ineffective performers. Lord's work is particularly noteworthy because of its systematic nature, and because of the compelling case that has been made for reliance upon implicit theories of leadership. In fact, to our knowledge, there has never been clearer evidence of the role of cognitive processes in any other type of evaluation. Nathan and Lord (1983) applied this approach to the study of appraisals and demonstrated that categorization processes can produce halo error, especially when ratings are based on information recalled from memory.

## Conclusions

There have been a number of cognitive models of the appraisal process that have guided and continue to guide appraisal research. Although there have been programs of research not clearly tied to any one particular model (which are discussed below), all of the cognitive research that has been conducted has been concerned with one or more of the processes in these models.

In reviewing the various models, several points become salient. One relates to the historical development of these models. As we noted, many of the ideas presented in the recent models have been proposed before, some over 30 years ago by Wherry (1952). Very little that is truly new is contained in the later models and, as has been suggested by critics, the field may be witnessing a variation of the notion that those who are ignorant of history are doomed to repeat its mistakes. Those interested in performance appraisal have not necessarily repeated mistakes, but do appear to have ignored a solid theoretical foundation for cognitively-oriented research on performance appraisal. On the other hand, more recent models have taken advantage of research done subsequent to Wherry's

paper to enrich and expand upon many of his ideas. Specifically, the emerging social cognition literature has allowed scholars to take principles of cognitive psychology as applied to the more general case of person perception, and use them to better understand the appraisal process. For example, Wherry's (1952) awareness of the role of schemata has been modified by more recent work on categorization to provide a more sophisticated view of rater's reliance upon preexisting configuration of traits and behaviors in making their evaluations. Furthermore, Wherry's (1952) concerns over memory and forgetting have been modified by more recent work on encoding and patterns of information organization in memory to provide a clearer picture of exactly how memory processes influence ratings.

Another interesting point about the various models is the extent to which they are really similar. All of the models begin by criticizing past research on simple rater training programs and rating instrument development and then propose a basic four-step process whereby raters acquire, store, recall, and then combine performance information to make judgments. And, of course, all of the models view the rater as an active collector or proposed intervention to improve ratings.

Where the models differ, generally, is in their emphasis on different model components. Far from being a problem, however, this may actually be advantageous to the field as a whole. If progress can be made on several fronts, it may soon be possible to step back, put the pieces together, and get a fairly complete picture of the appraisal process. The present paper will attempt to do this to some extent, but, as shall be seen, most of the published research has appeared in the past five years and there is, hopefully, much more to come. In any case, it is probably best to separate the research into two camps. One is interested primarily in information acquisition and storage in memory. Although this is the emphasis in the DeNisi et al., (1984) model, other researchers interested in memory processes (such as Kevin Murphy and his associates) do not necessarily base their research on that model. The other major camp emphasizes information processing, although here too there is interest in memory processes. Referring back to Ilgen and Feldman's (1983) models of automatic and controlled categorization processes, it may be possible to view these two approaches as differing in their concern with controlled processes (and thus information acquisition activities) versus automatic processes (which are less concerned with acquisition).

Thus, taken together, the proposed models do a reasonable job of explaining how and why raters purposefully acquire performance information; the importance of organization of information in memory for recall; the implications of the reliance upon both automatic and controlled processes; and the nature of categorization processes during appraisals. Of course, these are models, not empirical data. Nonetheless, looking upon these models as the beginning of understanding of these processes and assuming that these models will continue to guide empirical research, researchers may be well on their way to illuminating the role of cognitive processes in appraisals. The fact that a number of different scholars all

see the relevance of the same literature to appraisals, and all draw upon the literature to reach similar conclusions, differing only in the emphasis on different aspects of cognitive processes, indicates that the field may be approaching some consensus on important variables, questions to be asked, and methods to be employed. The study of cognitive processes in appraisals thus may be approaching the stage of being paradigmatic research.

## RESEARCH ON COGNITIVE PROCESSES

As in any new area of research, empirical support for the cognitive models has lagged behind the development of these models. Nonetheless, there have been a number of studies testing their propositions and, if recent trends are at all indicative of interest levels, there should be much more research in this area in the near future. In order to review the existing research, however, an organizing framework is needed. We will follow the basic cognitive processing model underlying the various models discussed and separate the studies into those dealing with acquisition of information; encoding of information, storage, and recall from memory (memory issues); and information processing.

### Information Acquisition

As noted earlier, the DeNisi et al. (1984) model pays particular attention to the process of information acquisition and much of the research associated with this model has focused on acquisition. Since most of these studies will be discussed in detail in a later section, they will be mentioned only briefly here, as they relate to the emerging knowledge base on cognitive processes. There have been other studies examining acquisition processes, however, and they will be discussed here. Two early attempts to study acquisition in appraisal contexts can be found in two dissertations carried out at Purdue University (DeNisi, 1977; Matte, 1981). Both of these studies concentrated on the amount of information sought, although information type was considered in terms of whether the information indicated good or poor performance. Unfortunately, the DeNisi (1977) study posed unrealistic constraints upon raters' choices and so, even though specific patterns of information acquisition were found, these patterns are not likely to be useful in other settings. The Matte (1981) study was able to overcome the constraints of the earlier study, but focused only upon amount of information sought, finding that this amount varied as a function of appraisal purpose.

Zedeck and Cascio (1982) approached the problem from a different perspective. Rather than studying acquisition directly, they examined the types of information subjects attended to in making appraisal decisions using a policy-capturing approach. They found that raters differentially attended to positive or negative



information concerning different job dimensions as a function of the perceived purpose for the appraisals. In fact, Zedeck and Cascio (1982) reported that these purpose effects overcame virtually all effects due to rater training in their study. DeNisi, Cafferty, Williams, Blencoe, and Meglino (1983) attempted to study acquisition more directly, from a attribution theory approach (Kelley, 1971). DeNisi et al. (1983) found evidence that raters search for distinctiveness- and consensus-type information about ratees (cf. Hansen, 1980). When raters were constrained in their search, however, the type of information sought had no analogy from the attribution research.

These studies supported the notion that information acquisition, either in terms of actual search or in terms of attention, influenced the appraisal process. Once this was established, however, two important questions remained: What factors affect acquisition activities, and what impact do these activities have on ratings?

The Zedeck and Cascio (1982) study suggested that appraisal purpose probably affects acquisition. A subsequent study by Williams, DeNisi, Blencoe, and Cafferty (1985) supported that position and, in addition, found that purpose had stronger effects on information acquisition than on utilization. Cardy and Kehoe (1984) argued that differential selective attention accounted for their finding that field independent raters were able to provide more accurate explanation for differences in acquisition activities. Finally, Balzer (1986) studied attention by having raters keep performance diaries when both initial impressions and the centrality of the rating exercise varied. He reported that raters were more likely to record (i.e., attend to) information inconsistent with initial impressions, and that initial impressions interacted with the centrality of the decision to influence the number of incidents recorded. Together, these studies demonstrate that information acquisition activities are influenced by appraisal purpose, initial impressions, and individual differences.

The relationship between acquisition activities and rating accuracy has also been established. Murphy, Garcia, Kerkar, Martin, and Balzer (1982) reported a relationship between the accuracy of observations and the accuracy of subsequent ratings based on those observations. These authors suggested, however, that this relationship is a complex one, with different aspects of observational accuracy important for different rating tasks. They also concluded that observational accuracy was not a generalized ability. Their results established the first link between acquisition activities and rating accuracy. A different test of this link was provided by Cafferty et al. (1986). They compared the accuracy of ratings provided by raters acquiring information in a person-blocked pattern (i.e., all information on one ratee before moving to the next ratee), a task-blocked pattern (i.e., all information about one task before moving to another task), and a mixed pattern (i.e., which changes ratees and tasks for each observation). They found that acquiring information in a task-blocked pattern produced ratings least prone to halo. These ratings were about as accurate overall as those produced by person-blocked acquisition, with the mixed pattern produc-

ing the least accurate ratings. Cafferty et al. (1986) also found that these differences in accuracy were attributable to different patterns of information organization in memory.

In conclusion, the research to date, though limited, has shown that patterns of information acquisition can be identified, that acquisition "strategies" or selective attention is influenced by appraisal purpose, overall impressions of ratees, and may even vary as a function of individual differences, such as field dependence/independence. In addition, these studies have shown that accurate ratings follow from accurate observation and that acquisition patterns themselves have an impact on rating accuracy, but that this impact may be due to the pattern of information storage in memory. Perhaps, then, the role of information acquisition may be simply one cue for raters to use for organizing information in memory. There may be other cues as well, any of which might help raters organize the performance information they acquire, even when that information is not acquired as the result of purposeful, or controlled, search.

#### Memory Issues: Encoding, Storage, and Retrieval

According to basic cognitive appraisal models, once performance information is acquired it passes on to what has been largely a "black-box" memory. Here, research is concerned with the way in which information is encoded, stored, and subsequently recalled from memory. There have been a number of recent studies which have been able to demonstrate the importance of these memory processes for appraisal decisions.

Kinicki and Lockwood (1985), for example, studied the initial encoding of information, specifically examining how performance information was segmented or unitized by raters. A rater observing a continuing stream of performance must find some way of dividing this flow into meaningful segments or units. Drawing upon the work of Newton (1973; Newton, Engquist, & Bois, 1977), Kinicki and Lockwood predicted differences in how performance information would be segmented as a function of appraisal purpose. Although they found no support for differences in the size of segments, they did find evidence for different patterns of segmentation as a function of purpose.

Other researchers have attempted to identify specific categories used by raters to encode information, or at least to demonstrate that raters rely upon such categories in their processing of performance information. Ostroff and Ilgen (1986) attempted to do both. Their study was noteworthy because of the effort they put into attempting to identify rating categories which might be used by raters. Although several hypotheses tested were not supported, Ostroff and Ilgen (1986) did find that raters who used categories consistent with those referenced in the rating instruments they used were more accurate in their ratings than were raters using inconsistent categories. The interesting implication here is that it may not be so important what categories raters use as long as they are consistent



with the demands of the rating context. This was also suggested by Ilgen and Feldman (1983), and, more recently, by Feldman (1986).

Another approach to identifying rating categories draws upon procedures more commonly used in the marketing area. Jolly, Reynolds, and Slocum (in press) employed procedures based on means-end theory (Gutman, 1982) to identify cognitive concepts salient to raters during appraisals. Specifically, a cognitive map was constructed for raters in the sample (i.e., nurses), consisting of attributes, consequences, and values, which was used to identify raters' perceptual orientations. The authors then demonstrated that raters' values were the most important rater input into the appraisal decision. These values dictated what was viewed as acceptable and unacceptable performance and were used by the raters to evaluate performance. Blencoe (1984) also examined rater categories of effective and ineffective performance, constructing prototypes of good and poor workers using procedures originally developed to assess stereotypes (McCauley & Stitt, 1978). He not only found that such prototypes existed, but also found that they affected ratings in much the same way as halo bias.

Banks and Roberson (1983) also found evidence of reliance upon categories for the processing of performance information. Their study deserves special note because of the effort devoted to understanding all aspects of cognitive processing engaged in by raters during the rating process. Their choice to study few raters in great detail makes research such as theirs costly, but it provides a richness of information not usually available. Nathan and Lord (1983) specifically set out to compare a traditional and a category-based processing model. The traditional model was seen as one where observed behaviors are integrated into independent dimensional schemata, while the categorization model was seen as one where observed behaviors are integrated into global categories and category prototypes are then used for making ratings. Nathan and Lord (1983) reported that the traditional model was sufficient for explaining both immediate and delayed ratings, but they also found a tendency to distort recall of performance incidents in a way consistent with general impressions, a finding that could only be accounted for by a categorization model. Although Nathan and Lord (1983) were more concerned with processing of information, their results clearly suggest that raters do rely upon categories when originally encoding performance information.

Several other studies, not directly concerned with memory models, are also of interest here. Cooper (1981b) and Murphy and Balzer (1986) both found greater evidence of halo in ratings that allowed the rater to draw upon broad categories of performance information stored in memory. The results of both studies indicate that information is stored in memory according to categories which tend to capture general impressions but obscure details. Murphy, Martin, and Garcia (1982) also reported that general impressions lead readers to recall specific details consistent with the general impression activated, lending further support to this position. In general, then, there is a considerable body of data to suggest that raters do possess identifiable categories that are used to encode and store

performance information. It is not clear what the domain of possible categories might be, but it does seem that raters are active processors of information in memory through the use of categories.

A final issue relating to the role of memory in appraisal decisions is concerned with retrieval from memory. As has been suggested earlier, recall can be affected by general impressions about ratees. Further, Murphy, Gannett, Herr, and Chen (1986) have suggested that the effects on ratings they found for subsequent performance information was due to biased recall. Clearly, accuracy in ratings is related to the accuracy of information retrieved from memory. A series of studies by Williams and his associates (Williams, 1984; Williams, DeNisi, Meglino, & Cafferty, 1986), however, has shown that the study of recall is further complicated by the fact that information acquired and encoded for one purpose must often be recalled and used for a different purpose. This phenomenon is known as the problem of reprocessing objectives (Anderson & Pichert, 1978) and suggests that information originally acquired for one purpose must be reprocessed before it be used for a different purpose. Both of the studies mentioned found that it was the initial or original purpose that dictated how information was organized, which would mean that reprocessing would tend to be difficult and could be a source of inaccuracy in ratings. This is especially significant if one assumes that any information acquired by means other than purposeful search (and even some acquired through such search) must be reprocessed before it can be used for performance information. This point will be discussed in more detail later. For the present, however, it is clear that memory processes play a critical role in appraisal decisions.

### Information Processing

The last broad class of cognitive activity in appraisals involves the integration and combination of information recalled from memory to form a judgment. Many of the studies in this area, which have been concerned with the reliance upon categories for processing information, were discussed in the previous section as evidence for the existence of such categories (e.g., Blencoe, 1984; Cooper, 1981b; Murphy & Balzer, 1986; Murphy, Martin, & Garcia, 1982; Nathan & Lord, 1983).

There have been other studies, however, which have examined additional issues related to information processing. Murphy and his associates reported that ratings are influenced by levels of prior performance (Murphy, Balzer, Lockhart, & Eisenmann, 1985), and that levels of performance subsequent to the target performance could influence ratings when those ratings were made one day after observation (Murphy, Ganett, Herr, & Chen, 1986). It seems that performance information, other than that which is to be rated, is integrated with the target information in order to form an evaluation. Although Williams et al. (1985) failed to find any effects for appraisal purpose on information processing (only

acquisition), Pulakos (1986) demonstrated different rating tasks required different types of training, including training for information processing, suggesting a potential role for purpose on processing strategy. Also, Zedeck and Cascio (1982) found that raters combined information differently when ratings were to be used for different purposes.

Recently, Lord and Alliger (1985) compared the validity of different information processing models of social perceptions. They found that perceptions of leadership were heavily influenced by the frequency with which leader behaviors occurred. Prototypicality of behavior was found to affect perceptions, but not to the same extent as behavioral frequency of the person being observed. One implication of this study is that perceptions of subordinate behavior are determined in the same manner; that is, ratings may be guided by frequency-based estimates of performance (cf. Kane & Lawler, 1979).

Perhaps one of the more interesting directions for research on processing activities concerns the role of rater affect. Affect, or the rater's personal feelings toward the ratee, as well as rater mood, have been proposed as an important component in other areas of social evaluation (Bower, 1981; Zajonc, 1980), and so it is reasonable that emotions might influence the processing of performance information as well. Cardy and Dobbins (1986) examined this possibility, investigating how "liking" influenced ratings of a teacher's performance. They reported that liking did influence the accuracy of teacher ratings, but only when liking varied across a set of ratees (as opposed to a constant level of liking). They interpreted this to mean that liking influenced the processing of performance information and was actually used as an input for the evaluation. This would make it extremely difficult to remove the biasing effects of liking from ratings. Williams and Keating (1987) demonstrated that affect influenced recall as well as ratings, further supporting the notion that affect biases processing of performance information. Thus, affect may influence ratings by influencing what is recalled or it may influence ratings independently but also influence recall. In any case, affect is a potentially interesting variable for future research.

### Conclusions

It is clear that research on cognitive processes in appraisal decisions has moved beyond the theorizing stage. The empirical research to date is not only noteworthy because it has demonstrated that cognitive processes play an important role in appraisal decisions, but also because of the volume and the systematic nature of the work itself. Studies have not only demonstrated that rater acquisition strategies exist, but also that these strategies can affect rating accuracy. Studies have shown that raters possess categories of ratee effectiveness, and often rely upon them in the rating process. Further, studies have shown the critical role of memory processes in appraisal decisions, and have demonstrated how a variety of factors can influence how performance information is processed. The field has

come a long way. Nonetheless, there are clear limitations to the research that has been conducted thus far, and there is the need for much more research in the future.

The most obvious limitation to the research to date has been its almost exclusive reliance upon laboratory studies involving "paper people" or videotaped performance. Although it has been argued that laboratory studies are fine for studying the judgment process, the rating process includes other contextual variables that can only be studied in field settings (Bernardin & Villanova, 1985). It is also noteworthy that paper people appraisal studies tend to produce larger effects than do field studies (Murphy, Herr, Lockhart, & Maquire, 1986), probably due to the absence of potentially biasing environmental cues in lab studies. These, and other, validity concerns will be discussed in more detail in the final section of this paper. A final point to be made here concerning future research deals with the need to demonstrate the practical implications of this line of research. This is, of course, the ultimate challenge to cognitive approaches to appraisal decisions. Researchers have been able to demonstrate with some success that cognitive processes affect appraisal decision, now they must demonstrate that cognitively oriented interventions can improve those decisions. We will discuss some of these possible interventions later, after first reviewing results of studies based on the DeNisi et al. (1984) model.

### A RESEARCH PROGRAM BASED ON THE DeNISI ET AL. MODEL

The research program undertaken with the DeNisi et al. (1984) model provides a specific direction for future research efforts in performance appraisal, and proposes a number of organizational interventions aimed at increasing the effectiveness of appraisals. This research is presented as an illustration of how cognitive approaches to performance appraisal are useful and how research might progress. It should not be interpreted as a definitive statement of the field. Undoubtedly, recommendations and interventions will be developed from other lines of research as well.

The DeNisi et al. (1984) model, as may be recalled, is similar to other cognitive models, except that a greater emphasis is placed on information acquisition processes. Performance appraisal is viewed as "an exercise in social perception and cognition embedded in an organizational context requiring both formal and implicit judgement" (p. 362). Drawing heavily from Wyer and Srull's (1981; Srull & Wyer, 1980) model of category accessibility, the performance appraisal task is seen as a series of stages: (1) the rater observes job-related behavior of the ratee, (2) a cognitive representation of the behavior is formed, (3) this representation of behavior is stored in memory, (4) stored information is retrieved for formal evaluation, (5) the retrieved information is reconsidered and

integrated with other available items, and (6) a formal evaluation is assigned to the ratee using a particular rating instrument. Within this framework, the rater is seen as an active seeker of relevant information. "Even when the rater encounters performance information passively, he or she still takes an active role in deciding which bits of information shall be attended to and how that information will be encoded for later use (cf. Bargh, 1982)" (p. 362). This approach raises questions of how raters choose to actively search for information needed for their appraisal tasks, and what information they attend to when they incidentally or passively acquire information. Related to both of these questions are the issues of how the rater organizes information and categorizes performance on memory and how these memory processes relate to judgments.

Two avenues of research that have followed from the DeNisi et al. (1984) model reflect the questions posed above. Both are reviewed here. The first reflects the model's emphasis on the rater's purposeful search for information and asks how information acquisition strategies might affect performance appraisals. Several studies have established a role for search strategies and have identified the pattern of information organization in memory as a key link in this process. Some of these studies have already been alluded to (e.g., DeNisi et al., 1983; Williams et al., 1985); they will be described here in more detail as they relate to specific interventions. The second line of research has addressed how raters reprocess performance information when evaluating performance subsequent to the initial acquisition of information. Situations in which raters encounter performance information incidentally, during the course of other activities rather than as the result of a purposeful search for information, are examined. A distinction is made in this research between "on-line" judgments (i.e., where appraisal is salient) and memory-based judgments (i.e., where appraisal is not salient). The main interest of these studies has been to identify memory organizational processes related to accurate and inaccurate ratings under varying conditions of appraisal salience.

In general, two major points emerge from these studies and are emphasized. First, the typical cognitive strategies used by raters are aimed at increasing cognitive efficiency and economy; they seem to serve data reduction functions, and may occur "automatically." The processing strategies used by raters in most settings simplify the tasks of attending to, storing, and retrieving performance information in order to make judgments of others. Specifically, raters seem to make use of worker categories to process information for upcoming appraisals, as suggested by Feldman (1981) and Lord (1985). Systematic distortions in ratings have followed from these data reduction processes and interventions aimed at mitigating these biases have been proposed. Second, different types of information processing seem to occur in different appraisal contexts and for different types of appraisal tasks. The manner in which information is processed has been related to appraisal purpose, appraisal salience, and rater characteristics. Thus, "boundary conditions" must be considered when examining the cognitive processes involved in appraisal decisions (cf. Bernardin & Villanova, 1986).

### The Effect of Information Acquisition Strategies

It should be noted that raters may acquire information either purposefully, as the result of a planned search for relevant information, or incidentally encountering performance information during the course of other activities. The incidental acquisition of information will be discussed in a subsequent section; the present discussion will focus on the active search for information.

The information acquisition strategy used by raters may influence the rating process in several ways: it may affect (1) the amount of information acquired by raters, (2) the type of information attended to by raters, or (3) the order in which information is acquired. Raters who have acquired more (or less) information, or who have attended to specific types of information, as a result of their search strategy, may make ratings dissimilar from raters employing alternative search strategies. Obviously, search will influence ratings only to the extent that raters use different strategies to acquire performance information. Thus, the first question suggested by the DeNisi et al. (1984) model was whether meaningful rater acquisition strategies could be identified. This involved first establishing that relevant information could be organized in different patterns and that these patterns were theoretically meaningful.

The strategy used to search for information is a method of organizing inputs, and to the extent that performance data can be organized in different ways, different search strategies should emerge. The nature of the information used for performance appraisals offers several organizational schemes. Typically, such information can be referred to as nonindividualized data (Pryor & Ostrom, 1981); multiple ratees can be compared along common attribute dimensions. Specifically, the performances of different individuals can be compared across similar or identical tasks. This can be contrasted with individuated data in which information about each target or ratee involves a unique set of attribute dimensions. The presence of nonindividuated information suggests a number of different strategies for organizing data. Organization may be guided by the salience of task categories (e.g., "What was the level of performance on each task?"), worker categories (e.g., "How did each worker do on the separate tasks?"), or performance categories (e.g., "What were the incidents of good performance and poor performance?"). Organizing one's search by worker categories would, for instance, be predicted by Feldman's (1981) model. The organization of individuated data, on the other hand, may be more uniform because of the absence of task or attribute categories; acquisition is more likely to be guided by person categories, the only category common to all information items. Thus, the presence of nonindividuated data should give rise to different types of search strategies.

It was also necessary to adopt a theoretical framework for categorizing raters' search strategies. DeNisi et al. (1984) suggested that the decision-making task facing a rater shares many key characteristics with the task of attributing some observed behavior to either a situational or dispositional cause. A rater observing

a ratee's performance on a certain task must try to decide if that incident is typical of that ratee and, therefore, could be expected in the future. If the performance could be attributed to an internal, stable cause, the rater could then consider this a true example of the ratee's performance (cf. Weiner, Frieze, Kukla, Reed, Rest, & Rosenbaum, 1971). This line of reasoning led to the utilization of attribution theory for categorizing types of performance information. According to this schema, a rater acquiring nonindividuated data could seek a distinctiveness type of information (i.e., the same worker's performance on other tasks), a consensus type of information (i.e., comparing other workers' performance on the same task), or a consistency type of information (i.e., the same worker's performance on the same task on another occasion). Drawing from the work of Hansen (1980), it was hypothesized that raters would prefer distinctiveness or consensus types of information when making appraisal decisions. Thus, they would structure their search around worker or task categories, respectively.

DeNisi et al. (1983) conducted a set of studies to see if these strategies were used by raters when nonindividuated data were made available to them. The first study examined acquisition strategies when no constraints were placed on the amount of information raters could request. One interesting finding from this study was that subjects requested limited amounts of information; the mean number of incidents viewed was less than half of the available information. The type of strategy used by raters was assessed by the order in which information was requested and by the frequency in which the rater's search changed ratee, task, and performance categories. Three predominant strategies were found; one where search was organized by ratees, one where search was organized by tasks, and one where search was organized by repeated task performances. When search was organized by ratees, subjects displayed a tendency to request information about how an individual performed across tasks before requesting information regarding another worker's performance (i.e., analogous to distinctiveness information in attribution theory research). This type of search was the most prevalent strategy, occurring for 44% of the raters. When search was organized by tasks, subjects displayed a tendency to search for information on how the different workers did on one task before searching for information regarding their performances on another task (i.e., analogous to consensus information in attribution research). Thirty percent of the raters employed this type of search. Finally, when search was organized by repeated performance, subjects displayed a tendency to search for how one individual performed a specific task on repeated occasions. This type of search is similar to a search for consistency information in attribution theory, and occurred 18% of the time. A small percentage (8%) of raters displayed none of these strategies but, instead, displayed a tendency to switch both persons and tasks on successive requests. The amount of information acquired under these different strategies did not differ reliably.

In study 2, DeNisi et al. (1983) constrained the number of items raters could acquire during their search. The preferences for the previously identified search

strategies changed in this experiment; nonblocked strategies resulted 57% of the time, task-blocked strategies resulted 34% of the time, and person-blocked strategies resulted 8% of the time. The search strategies used when no constraints are placed on raters appear to be quite different from those used under less-than-ideal, or constrained, situations. In addition, the most prevalent strategy in the constrained condition (i.e., the nonblocked search) does not fit with the adopted theoretical framework. Thus, attribution theory was not able to completely describe the range of search strategies undertaken by raters across different types of rating situations. It was subsequently decided to label the identified strategies according to the structure they imposed on the acquired information, rather than in attribution theory terms. That is, they were referred to as person-blocked, task-blocked, and nonblocked strategies.

Subsequent studies confirmed that raters prefer person-blocked acquisition strategies when not severely restricted in the extent to which they can search for information (Cafferty et al., 1986; Williams et al., 1985). It was also confirmed that raters sought less than half of the information available to them, and that no reliable differences in the amount of information obtained occurred for the different strategies.

### *Strategies for Different Appraisal Purposes*

The preference for person-blocked search strategies found in the above studies occurred when subjects were given a general performance appraisal task. That is, they were told that they would just have to rate the workers, with no particular purpose specified. It is possible that different appraisal purposes (e.g., administrative versus employee development purposes) may result in different types of strategies. Williams et al. (1985) tested this hypothesis by having subjects search for performance information for one of three appraisal purposes: promotion, salary allocation, or remedial training. The type of appraisal decision was also varied. Subjects were instructed to either rate each worker (i.e., referred to as "deservedness decisions" and perhaps similar to an absolute rating method) or to select one worker (i.e., referred to as "designation decisions" and perhaps similar to a relative rating method) for the specified purpose treatment. Overall, a preference for person-blocked search was found across purpose and decision type. There was, however, a higher level of task-blocked search among raters in the designation condition than the deservedness condition, especially for salary allocation. Thus, while person-blocked search strategies are generally favored, different strategies may occur for specific purposes.

### *Summary*

The initial work on acquisition strategies seems to suggest that raters have systematic strategies they prefer to use when gathering information about multi-

ple ratees. Overall, raters show a preference for person-blocked strategies when not constrained in their search. This type of strategy is consistent with the Feldman, Wherry, and Lord models, and may allow relatively efficient assessments of overall ability to be made. Raters also appear to be content with less than the majority of available information. Together, these processes may allow the rater to make inferences about ratees' performance without resulting in information overload from attending to all the available behaviors. A (seemingly) sufficient, albeit limited, amount of information about each worker's performance is processed in a fashion which allows the rater to form quick impressions of proficiency. A cognitive efficiency explanation seems to fit these data; enough information is gathered so that raters feel confident in their judgments, but their processing abilities are not taxed. Raters may also display cognitive efficiency in their search for information under time constraints. In these situations, raters seem to maximize person and task information by switching person and task categories frequently. They may be attempting to get as much information about each task and person as possible and use this information to generalize to other incidents.

One reason for the preference of person-blocked strategies may be that such strategies correspond to the raters' typical organization schema for encoding and storing information about others. The social cognition literature has suggested that when subjects receive information about multiple targets they naturally organize the information around the individual targets (Pryor, Ostrom, Dukerich, Mitchell, & Hernstein, 1983; Srull & Brand, 1983). In addition, Hamilton, Katz, and Lierer (1981) found that impression formation instructions led observers to organize information about target others in a person-blocked manner. Because the performance appraisal task may be considered similar to the impression formation task (i.e., both tasks require raters to form judgments of others along an affective dimension), the same method of organization in memory may be preferred for both tasks. While there has been much debate as to whether persons naturally organize social information around individuals (see the discussion in the following section), there still may be a tendency on the part of observers to group all attribute information about a person together during the judgment process. This strategy may allow one to make inferences about ability fairly quickly, or may activate existing schemata which guide further ability estimates. In either case, the need for further information is diminished.

The finding that raters in these different studies select only a portion of the information available to them (actually less than 50%) is also noteworthy. First of all, this finding is consistent with findings in the areas of consumer behavior and decision making where subjects have also been found to terminate their information search after acquiring only a portion of available information (e.g., Jacoby, Chestnut, Weigel, & Fisher, 1976). The relatively small amounts of information collected may be sufficient to activate existing rating schemata which,

in turn, fill in any missing detail for the raters (cf. Alba & Hasher, 1983). This undersampling of behavior was listed as one cause of illusory halo by Cooper (1981b). Additionally, raters may ignore repeated task performances for the same reasons observers ignore consistency or base-rate information when making attributions or predictions of others' behavior (Kahneman & Tversky, 1973; McArthur, 1972; Nisbett & Borgida, 1975). They choose, instead, to generalize from the specific observed incident to other performances on the same task. Thus, the search for limited amounts of information may trigger the use of schemata during the rating process. Typical errors in judgment resulting from the overreliance on schemata may be expected to follow.

The demonstration of different search strategies raises the questions as to how and whether the use of different strategies would influence ratings. It appears doubtful that search strategies influence ratings through the amount of information provided. There was no reliable difference in the amount of information provided by the strategies identified in our studies. It can be assumed, therefore, that the acquisition-rating relation is mediated by other factors. One possible factor is the way in which information is organized during encoding. Some raters, for instance, may encode information about one worker all at once, while other raters may encode information about one worker at separate occasions, as they move across task categories. Further, since the manner in which information is encountered affects the way it is organized in memory (Ostrom, Pryor, & Simpson, 1981; Pryor & Ostrom, 1981; Srull, 1983), it is possible that different appraisal search strategies generate different organization and storage networks in memory. This subjective organization of information, in turn, may affect rating accuracy. The studies reported below test the hypothesis that the organization of performance information in memory influences performance ratings.

### Subjective Organization of Performance Information

Observers subjectively organize data in memory by either fitting the data into a pre-existing framework or by creating new thematic categories which impart meaning to it (Loftus & Loftus, 1976). The retention and recall of material is superior when it is organized into meaningful categories than when it is unstructured (Bousfield, 1953; Mandler, 1967). Ostrom, Pryor, and Simpson (1981) found greater recall of person attributes when information was stored in memory by person categories than when it was not stored in any discernible manner. Srull (1983) also found greater recall of behaviors when information was stored in memory by person categories. In addition, the subjective organization of information in memory has been related to more accurate judgments of others (Hamilton et al., 1981). Thus, in the performance appraisal situation, raters may be better able to retain, recall, and evaluate performance information if they subjectively organize the information in memory. What has not been addressed is whether certain organizational strategies (e.g., person-blocked organization) lead to bet-

ter judgments than others (e.g., task-blocked organization). Research related to this issue will be discussed after first considering factors that influence the organization of information in memory.

Ostrom et al. (1981) found that the familiarity of target persons is related to the level of organization by person categories in memory; the more familiar targets are to the observer, the more likely that information will be organized in memory by persons. Temporal blocking of person information during acquisition has also been found to affect the organization of information in memory and recall. For instance, contiguous presentation of items related to individual targets facilitates clustering by persons compared to noncontiguous (nonblocked) presentation (Cofer, Bruce, & Reicher, 1966; Hernstein, Carroll, & Hayes, 1981). This finding is relevant to the patterns of information acquisition identified earlier. Specifically, it can be hypothesized that the acquisition strategies identified by DeNisi et al. (1983) would lead to corresponding organization strategies in memory; person-blocked acquisition should result in person-blocked organization in memory, task-blocked acquisition should result in task-blocked organization in memory, and nonblocked acquisition should result in no clear organization pattern in memory. The benefits of subjective organization would be expected to follow.

In addition to these contextual influences on organization, there may be natural tendencies to organize information according to specific categories. Such tendencies may be of particular importance when contextual cues are absent during processing. For example, under time constraints, raters gather information in a nonblocked manner. Since organization is absent from one's search in such situations, other factors must account for the presence, if any, of organization in memory. It has been suggested that subjects organize information automatically around target individuals. If raters naturally organize information about target individuals in memory around the individuals, then raters acquiring information in a nonblocked manner should show the same level of recall as raters acquiring information in a person-blocked manner. There has recently been much debate, however, in the person memory literature as to whether person organization occurs naturally. Early researchers in the area assumed that target persons formed the major nodes in observers' associative networks. Ostrom and his colleagues (McCann, Mitchell, Ostrom, & Hernstein, 1983; Ostrom, Pryor, & Simpson, 1981; Pryor & Ostrom, 1981; Pryor, Ostrom, Dukerich, Mitchell, & Hernstein, 1983), however, have recently presented evidence that social information is not naturally organized in memory around persons. In fact, they found little evidence for the automatic use of person categories in memory. Srull and Brand (1983) offered a resolution for this debate. They argued that search, encoding, and retrieval processes will depend on the initial processing objectives of subjects, and that subjects will not organize information around target individuals in memory when person category cues are weak. Pryor and Ostrom (1981), they claimed, examined situations in which subjects received memory set instructions, and thus person categories were not salient to them. When target individuals are

salient to processing objectives, person categorization in memory should occur. Srull and Brand (1983) found support for this hypothesis. They instructed subjects to study information about target-others, either in preparation for a recall test or an upcoming interaction with the target. Subjects who were given the anticipated interaction processing objective "were able to mentally separate or compartmentalize the information associated with each of the individual targets" (p. 223). These subjects also showed higher levels of recall than memory set subjects. The anticipated interaction with the target presumably increased the strength of person category cues, thus allowing person organization in memory to occur. Since performance appraisal processing objectives make person categories salient to raters, it may be expected that person organization would naturally occur in such settings.

Thus, raters may use either contextual cues, such as the organization inherent in their search strategies, or specific category cues, such as individual targets, to organize information in memory. Recall of performance information, and presumably ratings, should be more accurate when information is organized in memory than when it is not organized. The following studies tested these notions.

Cafferty et al. (1986, Study 2) found support for the hypothesis that organization in memory is influenced by contextual factors. Person-blocked presentation of information led to person-blocked organization in memory, task-blocked presentation led to task-blocked organization in memory, and nonblocked presentation led to no discernible organization in memory. Thus, the contiguous encoding of items during information acquisition is likely to influence the manner in which information is organized and stored in memory. It is also important to note that there was not a predilection on the part of raters toward organizing nonindividuated data in memory by persons. This seems true for inexperienced raters, at least. Perhaps experienced raters may be more likely to reorganize nonblocked information into person categories. Additionally, the performance appraisal task itself may not have been salient enough to raters (subjects were told to form impressions of the organization and its workers) and therefore person category cues may have been weak relative to context cues. At any rate, these results established the link between search strategies and subjective organization in memory.

It was also predicted that greater levels of organization in memory would result in greater recall and more accurate ratings. Raters organizing information by either persons or tasks (i.e., the person- and task-blocked presentation groups), indeed, had more accurate recall and were able to provide more accurate ratings than raters who did not organize information in memory (i.e., the nonblocked presentation group). There was no significant difference between the person- and task-blocking groups in terms of recall accuracy. In addition, no significant differences were found between the groups in terms of the overall ratings assigned to the ratees. In terms of task ratings, however, raters in the task-blocked group were less influenced by the ratees' overall proficiency and were better able to differentiate among varying levels of performance exhibited by the same ratee.



Specifically, they rated poor performances by a worker significantly lower than that worker's good performances, while raters in the person-blocked or nonblocked group did not significantly differ in their ratings of good and poor performances by the same worker.

These results demonstrate that information acquisition patterns may influence the subjective organization of performance data in memory. They also establish a link between subjective organization in memory, recall, and rating accuracy. Organizing information in memory according to performance categories led to higher levels of recall and resulted in more accurate ratings than did the absence of subjective organization. Therefore, more performance information was accessible to raters in the task- and person-blocked conditions than in the nonblocked condition. It is not surprising that segmenting performance information in memory according to each individual worker increases the accessibility of performance incidents. Once a particular person category in memory is accessed, the performance associated with that person should be easily retrieved because of the proximity of performance items to one another. Organizing information by task categories, however, appears to render information as accessible to raters as the use of person categories. Although the same performance information is stored in memory, it is just structured differently, and task dimensions seem to be as effective retrieval cues as individual workers.

Person- and task-blocked organization may, however, lead to differences in ratings. Person-blocking was found to decrease intra-ratee discriminability in the Cafferty et al. (1986) study. Perhaps the organization of information in memory by persons has other consequences besides increased accessibility of information. Impression formation has been associated with the organization of information by persons (Hamilton et al., 1981). A cognitive operation that follows from impression formation is abstraction (Anderson & Hubert, 1963). Schul (1983) defined abstraction as "the process by which a global meaning, gist, or summary is abstracted from the communication and stored in memory" (p. 46). This abstract, global impression of the individual may be stored separately in memory from the specific behavioral details upon which it is based (Anderson & Hubert, 1963; Ebbeson & Allen, 1979; Wyer, Srull, & Gordon, 1984). Two different codes, an impressionistic code and a detailed behavioral code, may therefore represent person information in memory. The impressionistic code may be more salient when person-blocking occurs. When asked to make judgments about an individual, raters may automatically access the global impression and use this as a basis for their evaluations (cf. Nathan & Lord, 1983). This has important implications for rater bias in performance appraisals. If person categorization and impression abstraction decrease the discriminability of specific behaviors, then raters organizing information in memory by persons may be prone to halo errors. A rater who has abstracted a favorable impression of a worker may be likely to rate that worker as performing well on all tasks, regardless of actual behavior. Thus, the greater accessibility of information that occurs with person categoriza-

tion in memory may be offset in a judgment situation by the abstraction of global impressions which are likely to influence ratings. An effective rating intervention would be one that prevents abstraction processes while maintaining the accessibility of behavioral information. That is, raters may be encouraged to use behavioral codes rather than impressionistic codes. Such an intervention would be similar to Lord's (1985) suggestion that automatic processing be replaced with controlled processing on the part of the rater.

The organization of information in memory by tasks would appear not to result in abstraction processes, or at least make it less likely that impressionistic codes are formed and stored in memory. This would imply that if the use of task categories in memory results in the same level of information accessibility as person categorization, then ratings may actually be more accurate since halo bias should decrease. The results from the Cafferty et al. (1986) study may be interpreted as supporting this notion, although further research is needed to substantiate this claim.

### Summary of Subjective Organization Processes

These studies identify the subjective organization of information in memory as having a key influence on performance evaluations. Figure 1 presents a model of this process. Drawing from the DeNisi et al. (1984) model, ratings are based on the conceptualization of performance in memory. The manner in which one subjectively organizes information is a key determinant of how one conceptualizes ratee performance. Different organizational strategies are likely to result in different representations of performance (e.g., the use of person categories in organization increases the likelihood of abstraction and impression formation

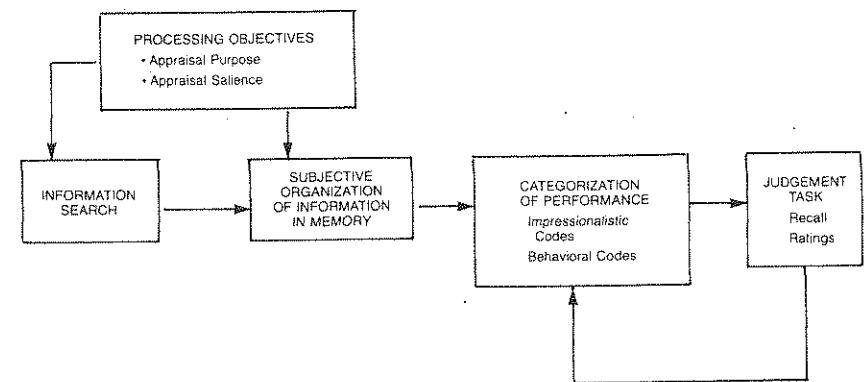


Figure 1. Model of subjective organization processes.



tendencies). Rating bias is a function of the representations that follow from one's organizational processes. If strong impressionistic codes are established for ratees, behavioral or episodic codes are often ignored or de-emphasized during the rating task. Distortions in ratings, such as halo or leniency, are likely to occur when global impressions mask behaviors. Information search strategies exert their primary influence on the rating process through their effect on the subjective organization of information.

The link between the subjective organization of information and judgments is expanded in Figure 1 to take into account (1) the effect of previous appraisals on subsequent information processing, and (2) the influence of processing objectives on the judgment process. The feedback loop from the judgment stage to the conceptualization and search stages reflects the view of the appraisal process as a dynamic rather than a static process. It is hypothesized that once decisions are made regarding a ratee's performance, the conceptualization of behavior in memory, as well as the subsequent search for information, are altered to reflect those decisions. The processing objectives box in Figure 1 reflects the importance of encoding operations for subsequent information organization and judgment processes (Srull & Brand, 1983). Together, these two variables (i.e., the effect of previous judgments and appraisal salience) represent the second line of research within the DeNisi et al. (1984) model: How do raters reprocess performance information when they must rely on information already stored in memory to make their evaluations?

### Reprocessing: The Effect of Previous Judgments

A recent study by Williams et al. (1986) supported the link between judgments and conceptualizations of performance in memory in Figure 1. Subjects acquired information initially to make either a "designation" decision (i.e., to pick one of four workers as the most deserving of contracting work, based on merit) or "deservedness" decisions (i.e., rate each worker as to his deservedness of this merit work). Subjects acquired information in either a person- or task-blocked manner, and the ratees displayed three levels of work proficiency (i.e., below average, average, above average). Two days later when asked to rate the workers from memory, subjects who initially made designation decisions could not discriminate between average and below average workers. It is argued that these raters had initially formed a dichotomous representation of performance (i.e., the best worker versus all others) and stored this representation in memory. Later, when asked to rate all the workers, they could not distinguish among the nonselected workers, except that they were poorer performers than the best worker. Subjects who made initial deservedness decisions were only able to differentiate among all proficiency levels if they acquired information in a person-blocked manner. Perhaps the use of person categories allows raters to form accurate representations of overall proficiency, and the nature of deservedness

decisions leads raters to compartmentalize information in memory around each individual worker, rather than groups of workers. If this is indeed the case, then these results seem at odds with the earlier notion that task-blocking would lead to greater rating accuracy (cf. Cafferty et al., 1986). The results of Cafferty et al. (1986), however, relate more to intra-ratee, rather than inter-ratee, discriminability. While the use of person categories may increase impression formation tendencies and reduce intra-ratee variability, it may increase the rater's ability to assess the overall proficiency of each worker. This line of reasoning suggests that different types of organization may be best for different types of evaluations.

This study also demonstrated cognitive economy on the part of the rater, as well as the tendency to automatically process performance information according to one's stored representation of performance. First, raters' representations of performance appeared to match the demands of their initial decision. For raters having to identify only the best worker, a simple compartmentalization of "the best worker" and "all others" may be the most parsimonious organization. Given the immediacy of the designation task, it is not necessary for raters to establish separate "compartments" or memory "bins" for each ratee; to do so may disrupt or delay focusing on the task at hand. Thus, a very general impressionistic code is established in memory, one in which behavioral detail is obscured. Raters also appear to automatically access the representation of ratee behavior they have stored in memory when faced with subsequent judgement tasks (cf. Lingle & Ostrom, 1979; Wyer, Srull, & Gordon, 1984). Raters who have constructed simple conceptualizations of performance in memory (i.e., a dichotomous representation of the best versus all other) may find it difficult to later make fine distinctions in their performance ratings. Again, the key to accurate ratings may lie in circumventing the rater's impressionistic codes and making his or her behavioral codes more accessible.

### Reprocessing: The Effect of Appraisal Salience

The view of the appraisal process presented up to now is based on studies in which the rating task was salient to raters. Raters initially acquired and processed information with the intent to rate the workers at some point in time for a certain purpose. This is not always the case in organizational settings. Raters have other responsibilities in addition to performance appraisal, and thus the evaluation of workers is not always their primary concern when performance information is encountered. They may observe behavior of subordinates while performing various job functions or duties, but only relate that behavior to performance appraisal standards later when formal evaluations are required. The processing of performance information may be quite different when appraisal is salient to observers than when it is not salient. For instance, the studies reported above suggest that raters make use of person categories to encode and store information about others for upcoming appraisal decisions. But the same categorization may not occur

when the appraisal task is not salient to raters. Recent research indicates that observers do not automatically make use of person categories to organize information in memory (Cafferty et al., 1986; Pryor & Ostrom, 1981). In addition, Srull and Brand (1983) have suggested that person organization in memory would be likely to occur only when person category cues are salient. Thus, when appraisal salience is low and person category cues are likely to be weak, the level of person categorization may be low, or at least not as high as when appraisals are expected. Differences in the recall and evaluation of ratee performance may follow.

To date, only a few studies have examined the appraisal salience issue. Barnes-Farrell and Couture (1983) found limited effects for appraisal salience on the recall of ratee behavior. However, it appears that the subjects in their low-salience condition were not given a competing processing objective and thus may not have been able to attend to performance information to the same extent as subjects in the high salience condition. Balzer (1986) found that the centrality of an appraisal task (i.e., salience) interacted with impressions of ratees to influence attention and encoding processes.

Williams (1984) provided a direct test of the effects of appraisal salience by assigning subjects either a performance appraisal task or a competing nonappraisal task. Performance was observed in a nonblocked fashion and all subjects were later asked to recall and rate the performances of the workers. The order in which raters performed the recall and rating tasks was counterbalanced within all experimental conditions. The level of person organization in memory, the accuracy of free recall, and rating accuracy were assessed. Overall, the level of person organization in memory was higher for high-appraisal salience than for low salience, lending further support to the hypothesis that raters use person categories to aid the processing of performance information, at least when person cues are relatively strong. In terms of recall, high-appraisal salience resulted in significantly more performance episodes recalled than low salience. Interestingly, high-appraisal salience also resulted in significantly more intrusions (i.e., incorrect matches between targets and performances) than low salience. It appears that raters in the high-salience condition adopted a looser response set than other raters. Having selectively attended to performance information, they may have been more willing to accept and report items retrieved from memory, including those which they were unsure of. In terms of ratings, limited effects were found for appraisal salience. High salience resulted in slightly more accurate ratings, but neither condition evidenced consistent differentiation between proficiency levels. When the relationship between recall and rating was examined, striking differences were found. Recalled performances and performance ratings were not significantly correlated when appraisal salience was high, but significant correlations were found when salience was low. These findings support the hypothesis that person categorization results in two codes being established in memory; a behavioral code in which specific behaviors are recorded, and an impressionistic code in

which a global evaluation or impression of the person is recorded (and updated by incoming information). The behavioral code guides the retrieval of items from memory, but the impressionistic code guides the judgment process. When appraisal salience is low, person categorization is less likely to occur (i.e., there was no evidence of person organization in memory), and ratings are more likely to be based on recalled information. Thus, when expecting to rate workers, raters may selectively attend to performance information, establishing both behavioral and impressionistic codes in memory, thereby being able to recall more specific episodes than when they are not expecting to rate workers. Accuracy in recall is also likely to be higher, but it may be limited by the tendency to "fill in the gaps" with impression or schema-consistent information.

Another set of findings from the Williams (1984) study involved the order in which subjects recalled and rated behaviors. Significant Salience x Order interactions occurred for the amount of recall, correct recall, and rating accuracy. When performance appraisal was not salient during observation, raters who performed the recall task prior to the rating task scored higher on the recall and rating measures than their counterparts who performed the rating task first. In fact, there were no significant differences on any of the measures between the raters in the low-salience/recall-first condition and the raters in the high-salience condition. Apparently, performing a lengthy search of memory and having to report the retrieved items during a reprocessing task increases the accessibility of items stored in memory relevant to the new task. An increase in rating accuracy can be expected to follow from the greater recall of relevant items, since ratings are based on recall under conditions of low salience.

This suggests that different processes may occur when appraisal salience is high than when it is low. Expecting to rate workers leads to the use of person categories in memory, impression formation, and low correlations between recalled behaviors and ratings. Not expecting to rate workers results in the use of other categories in memory (i.e., relevant to the task at hand), reduces the strength of impressionistic codes, and increases the relationship between recall and ratings. These processes are similar to the distinction between "on-line" judgments and "memory-based" judgments (Hastie & Park, 1986). On-line processing occurs when observers encode information with one primary purpose in mind. When a person judgment is being made, observers establish a representation of behavior in memory related to that judgment. This representation of behavior is continuously revised or updated as new information is encountered. When performance appraisal is salient to raters, they form an initial representation of each ratee's performance, and then adjust these impressions as new information is encountered. Hastie and Park (1986) reviewed studies in which subjects were asked to make on-line judgments and concluded that little or no relationship exists between recall and judgment in such situations. Rather, these types of decisions are based on impressions of targets that have been formed and stored in memory. The reliance on stored impressions may occur automatically, as suggested by Feldman (1981)

and Lord (1985). Memory-based judgments, on the other hand, are tasks in which decision makers must rely on the actual behavioral codes stored in memory, because impressions related to the required judgments do not exist or are weak. Such judgments occur when observers are not expecting to make the type of decision asked of them, as in the low appraisal salience situation. Hastie and Parks (1986) concluded that recall and ratings are related under these conditions.

Distortions in on-line judgments will occur when stored impressions do not fully capture the behavioral detail that exists, and raters are not encouraged to access behavioral codes in memory. Distortions in memory-based judgments will occur when relevant information is inaccessible due to memory failure. Williams (1984) suggested that structuring the reprocessing task increases the accessibility of information for memory-based judgments, and thus may be an effective means of increasing accuracy. Increasing accuracy for on-line judgments may be more difficult. One way to decrease bias for on-line judgments would be to circumvent global impressions in memory. That is, if raters could be encouraged to address behavioral codes instead of global impressions, ratings may more fully represent actual performance. This may be akin to processing information in a controlled, rather than automatic, manner.

Williams, Wickert, and Peters (1985) hypothesized that since person categorization is associated with impression formation during on-line judgments, having raters initially organize information according to a different categorization scheme, such as task-blocking, would increase the controlled processing of information and decrease rating distortions. They presented information to subjects under conditions of high and low appraisal salience, and instructed subjects to organize information in memory according to person or task categories. Task organization instructions did reduce person categorization in memory, although high levels of task categorization instructions led to greater within-ratee discriminability in ratings than did person organization instructions. It was argued that the task organization set disrupted the raters' naturally tendency to form impressions during on-line judgments and therefore induces raters to attend to actual behaviors.

#### *Summary of Reprocessing Studies*

Together, these studies show that attentional, encoding, and retrieval processes are affected by the saliency of the rating task, and that structuring or restructuring these processes may increase accuracy, especially when salience is low. Future research should address exactly how to structure these processes.

Two types of processing appear to occur during performance appraisal tasks. The first type of processing, which seems to be most prevalent, occurs when on-line judgments are made. Here, raters tend to use person categories to organize information and guide their recall and rating of behavior. Global impressions of workers may be formed which obscure behavioral detail. The second type of processing occurs when person category cues are not salient to raters during

observation and memory-based judgments are required. Here, raters organize information according to their initial processing objectives and later must reprocess information from an appraisal perspective. Rating bias is likely to occur in both processing modes. Person categorization results in greater correspondence between impressions and ratings than recall and ratings. As a result, perceived intra-ratee variability is likely to decrease and halo and/or leniency bias may result. Bias occurs for memory-based judgments due to inaccessibility of relevant items in memory, or from over reliance on easily recalled items. Once the nature and source of this bias is understood, interventions can be planned. Circumventing global impressions during on-line judgments and structuring the reprocessing task during memory-based judgments have been suggested as possible methods to increase accuracy. Since the organization of information plays a significant role in either process (see Figure 1), it may be possible to impose a general organizational schema that will increase accuracy in both situations. In order to do this, however, research must first establish which type of organization is most conducive to accurate ratings. Preliminary studies have addressed this issue.

#### *Interventions Based on the Organization of Information in Memory*

The outline presented in Figure 1 clearly suggests an intervention aimed at increasing rating accuracy; train raters to organize information in memory according to performance-relevant categories, in this case, either ratee or task categories. It would appear that contextual cues are effective ways of influencing the subjective organization of information. For instance, the contiguous pattern in which information is acquired will affect the organization of nonindividuated data in memory (Cafferty et al., 1986). But what about information that is not acquired in a contiguous pattern, such as when time constraints are present during one's search strategy or information which is incidentally acquired? It may be possible to provide methods by which raters can structure information in an easily accessible and efficient manner, or to identify other contextual variables which can act as cues to effective organization. Rater diaries are an example of the first strategy, while making the organization structure in memory compatible with the rating scales used during evaluation is an example of the second strategy. Both of these strategies have been examined recently.

Rater diaries have been advocated as a method of reducing the memory demands typically placed on raters (Bernardin & Walter, 1977). By documenting ratee performance, raters do not have to rely solely on their memory for performance appraisals. It is assumed that rating accuracy will increase since memory loss, a key contributor to inaccuracy, will be less of a factor.

Structuring performance diaries according to appraisal relevant categories may be an effective information processing tool. An unorganized diary, like the absence of subjective organization in memory, may reduce the accessibility of

relevant information to raters, while a diary organized by ratees or tasks may facilitate attention to relevant information. In addition, there may be some advantages of one type of organization over the other. Task-blocking may reduce general impression formation, and decrease the likelihood of halo bias. Person-blocking, however, may make it easier for raters to form a complete representation of a ratee's performance, since all information about that worker is consolidated in one section. A recent study by DeNisi, Gardner, and Cafferty (1987) investigated the effectiveness of different types of diaries on raters' organization of information and rating accuracy.

DeNisi et al. (1987) instructed raters to record performances in one of four diary conditions: (1) a free diary condition, in which raters were instructed to record ratee performances in any manner they chose; (2) a person-blocked condition, in which raters were given a diary divided into sections representing ratees and were instructed to record the performances of each ratee in the corresponding section of the diary; (3) a task-blocked condition, in which raters were given a diary divided into sections representing the task dimensions on which ratees were to be evaluated and were instructed to record the performances of each ratee in the diary according to the behavior dimension being observed; and (4) a no-diary control condition in which raters were not given a diary in which to record ratee behavior. Raters were also required to perform a parallel, in-basket task. This task added to the external validity of the study, and made it more critical that the raters organize information in some manner. After completing the observation and in-basket tasks, subjects were asked to recall and rate the behaviors of the (three) workers. The recall task was made without the benefit of the diaries and was used to measure the organization of information in memory, as well as information accessibility. Raters were able to use their diaries for the rating task. Results showed that subjects who did not keep diaries performed poorly on the recall and rating tasks. Not only did these subjects do worse than the other subjects, but they were actually unable to distinguish among the varying levels of proficiency exhibited by the ratees. They failed to make either inter- or intra-ratee distinctions in their ratings. It appeared that the use of diaries enabled raters to organize information in the face of competing demands, and increased the accessibility of performance information during the judgment tasks.

In the experimental conditions, the raters' subjective organization in recall followed from the organization of diaries. Person-blocked diaries resulted in person-blocked organization, and task-blocked diaries resulted in task-blocked organization. The vast majority of subjects in the free-diary condition chose to keep person-organized diaries, and their recall reflected this structure. A comparison of the three diary conditions indicated that person-blocked diaries led to more accurate recall and slightly better ratings. The major finding was that person diaries reduced errors in recall. That is, the amount of recall was similar, but these raters had fewer instructions and thus had higher memory improvement scores. In terms of ratings, the only finding was that person diary subjects were

able to distinguish among the different levels of performance within all ratees, whereas the task and free-diary subjects were able to do this for only two of the three ratees.

A second line of interventions suggested above involves the use of rating scales as organizing devices. Ilgen and Feldman (1983) have suggested that rating accuracy will be high when information is organized in memory according to the same categories used on the rating scales, provided that the rating scales themselves are psychometrically sound. Wherry (1952) suggested that task-blocked ratings decrease halo bias. While research testing this notion is mixed (Johnson & Vidulich, 1956; Johnson, 1963), no one has tested Ilgen and Feldman's compatibility notion. Perhaps, if raters know they will be using a task-blocked rating form and are able to organize information in memory according to task categories, behavioral information will be more accessible to raters and accuracy will increase. DeNisi and Summers (1986) provided subjects with rating scales organized around person or trait categories either before or after observation of performance. They found that providing raters with the scales before observation increased the level of organization in memory and also increased accuracy. Task organization, however, was not found to be superior to person organization. The fact that organization in memory was accomplished, and that it was compatible with the rating scales used, was the main determinant of accuracy. These results converge with those of Pulakos (1986), who found that rating accuracy was greater when rater training and rating scale format were congruent than incongruent. In addition, they support Feldman's (1986) hypothesis that the nature of the rating task and rating scales themselves can provide the structure for organizing and interpreting ratee behavior.

The results from the Williams' (1984) study also imply that information can be organized in a manner conducive to accurate ratings at the reprocessing stage. Perhaps imposing person or task organization or congruency between recall and rating format during a reprocessing task will improve the accessibility of information further. Attempts can also be made to reduce the data simplification that occurs during many appraisal decisions. Judgment tasks, such as designation decisions (Williams et al., 1986), which will result in simple conceptualizations of performance in memory (e.g., the best worker versus all others), should be discouraged and restructured so that more detailed representations of behavior are established. In the Williams et al. (1986) study, initial deservedness tasks resulted in more accurate post observation ratings than designation tasks, presumably because every worker's performance was represented separately in memory. Further research should examine how other appraisal tasks affect the representation of behavior in memory, and decisions which lead to detailed representations should be identified.

## CONCLUSIONS AND FUTURE DIRECTIONS

The models outlined in the opening sections identified the processing stages of the appraisal task and suggested possible sources of bias. The empirical evidence reviewed supports many of the models' propositions. Attentional, encoding, and retrieval processes engaged in by the rater reflect the need to process relatively large amounts of information in a busy environment in order to reach cognitively complex judgments. Rating inaccuracies can be seen as shortcomings in the rater's attempts to meet this challenge.

A number of conclusions can be reached relating the rater's processing of information and rating accuracy. First, the way in which raters choose to subjectively organize information in memory clearly affects rating accuracy. Differences in the organization of information in memory have been reliably related to changes in the amount of information recalled, the accuracy of recall, and rating accuracy. The subjective organization of information, in turn, is partially a function of information acquisition and encoding strategies. It may be possible to alter these strategies to obtain desired levels or types of organization in memory. Further research should address this issue.

While memory organization has been identified as crucial to accuracy issues, it is unclear what, if any, is the best type of organization. Cafferty et al.'s (1986) study might suggest task-blocking, whereas the Williams et al. (1986) and DeNisi et al. (1987) studies would suggest person-blocking. Perhaps there is no one method that is universally superior; moderating factors, such as appraisal purpose (Feldman, 1986) or individual difference variables may exist. Another possibility, which relates to generalizability notions, is that different organizational schemes than those considered here exist in more natural settings, and that these schemes may prove superior to those considered here. Nonetheless, imposing organization on performance data stored in memory should increase accuracy. Bias would be predicted to occur when raters do not possess performance-relevant schemata in memory.

A second conclusion to be reached is that different modes of information processing occur during rating tasks. These modes of processing may, in fact, occur as a function of subjective organization of behavioral information. Evidence has been presented showing that both impressionistic and behavioral codes are established by raters. Consistent with the person perception literature, raters display a strong tendency to rely on the impressions formed of others when making judgments. A natural result of this tendency is halo bias. Certain conditions, such as low familiarity with the task or ratees and less than optimal attentional and retrieval conditions, increase the likelihood of halo bias emanating from impressionistic codes. While halo bias may be less likely when using behavioral codes, different types of bias may occur. Overreliance on easily recalled or available items (Kahneman & Tversky, 1973), generalizing from specific instances and ignoring base rate information (Nisbett & Borgida, 1975),

or distorted relative frequency estimates (Lord & Alliger, 1985) may lead to rating distortions. (It could be argued, however, that behavioral frequency estimates are evidence for the use of impressionistic codes, especially given the authors' view of this type of processing as automatic.) To reduce bias resulting from impressionistic codes, interventions would have to make the congruency (or incongruency) between behavioral codes and impressionistic codes clear to raters; to reduce bias resulting from behavioral codes would involve increasing the accessibility of representative samples of behavior.

Researchers have also related automatic and controlled processing modes to rating bias, although, as suggested earlier, the relevancy and meaning of these terms need to be made clear. Most researchers would agree that automatic processing is likely to result in more bias, largely due to the use of faulty rating schemata (Lord, 1985). This prediction, however, needs to be directly addressed in future research. When each mode of processing is likely to take precedence over the other may be a more pressing issue. Answering this question may ultimately enable researchers to relate processing modes directly to bias. The following distinction is proposed for initial research attempts.

For many on-line judgments, automaticity should operate. Person categorization and abstraction of impressions of others should occur independent of the rater's control and in parallel when the judgment task is salient and familiar to the rater. Perhaps, behavior frequency estimates and conceptual similarity perceptions will develop as well as a result of these processes. Low correlations between recall and ratings should occur, and distortions in ratings would be related to inaccurate or underdeveloped impressions. For memory-based judgments, controlled processes would be more likely to occur. These types of judgments are to be expected when the appraisal task is not salient during encoding, or unexpected or unfamiliar circumstances arise. Retrieval failure and inaccessible information would be the major obstacles to rating bias in these situations.

A number of interventions have been suggested by recent cognitive research. Performance diaries have been suggested as one way of making actual behavioral information more accessible. Written records of performance may force the use of behavioral codes in addition to or in place of impressionistic codes. They would also increase the accessibility of information because memory retrieval failure would be reduced. The use of diaries, however, makes observational and attentional skills crucial. This area of performance appraisal has been under-researched, and thus steps should be taken to ensure that bias in observation and attention do not mitigate the advantage of using diaries. Another problem with the use of diaries relates to the issue of compliance. Many raters may not have the time to record all performances or may not see the utility of their diary. A recent study by Williams (1987) suggested this may indeed be likely.

Increasing the compatibility of the rater's subjective organization of information, memory retrieval patterns, and rating scale formats has also been suggested as a method for increasing accuracy. The assumption made by this approach is that by

making information input and output processes compatible, distortions in the transmission of information would decrease (cf. Feldman, 1986; Pulakos, 1986). This approach is similar to frame-of-reference training (e.g., Bernardin & Buckley, 1981), which attempts to provide raters with an organizing framework consistent with rating dimensions. Future research within this compatibility framework should help refine these techniques.

### Generalizability Concerns

Research on the cognitive processes involved in performance appraisal has been summarized and future research propositions have been offered. What needs to be considered before continuing along these lines, however, is the generalizability of the presented research. The immediate concern that arises from a perusal of this work deals with the external validity of studies using artificial stimuli and nonprofessional raters. Murphy et al. (1986) have addressed the issue of paper-people designs, and found larger effect-sizes for paper people than live or taped subjects. Fortunately, most of the research described here has made use of videotapes or real people as ratee stimuli. Banks and Murphy (1985) have also raised a number of related issues which need to be considered before pursuing research along the lines suggested. For instance, they have argued that raters in the typical laboratory experiment have not had to differentiate between information that is relevant to appraisals and information that is irrelevant. This relates to the laboratory's inability to capture the "busy" environment of raters in organizations. While recent studies have incorporated competing demands on raters into their design (e.g., DeNisi et al., 1987), future research should be planned with the intent of incorporating relevant and irrelevant information.

Other generalizability concerns arise regarding the theoretical assumptions being made by researchers in this area. It was mentioned earlier that the present models are able to expand on Wherry's earlier model because of recent advances in the social cognition area. Ilgen and Favaro (1985), however, raised a number of important concerns regarding the appropriateness of applying social cognition principles to performance appraisals. In particular, they questioned the absence of concern for the timing of judgments, consequences of behavior, and interactions between rater and ratees. These criteria could easily be built into future experiments, whether in the field or in the laboratory.

In an attempt to help researchers increase the external validity of their results, Bernardin and Villanova (1986) have proposed a "modal criterion setting." This is a practical model of the performance appraisal process, based on survey responses from practitioners, which specifies the respondents, the work setting, and appraisal conditions which are modal or characteristic of most appraisal settings. Research findings and interventions are considered relevant or generalizable to the extent that the studies contain all or most of the elements of the modal setting. The modal setting proposed by Bernardin and Villanova (1986) includes

the following: (1) the rater evaluates multiple ratees, (2) ratings are used for important personnel decisions, (3) raters feel they don't have enough time to accurately rate workers, (4) ratees sign off on the rating form, (5) ratees are aware of how fellow ratees are evaluated, (6) ratings are done on an annual basis, (7) raters are responsible for subordinate work products; (8) raters have performed the tasks being evaluated, (9) raters are confident in their ratings, (10) raters feel uncomfortable appraising others, and (11) raters evaluate traits. A comparison of these characteristics with those of the studies presented in this paper shows a fairly large degree of overlap. Eight of the modal characteristics have been included in the studies upon which the previous recommendations and interventions are based. Among those that have been consistently included are the use of multiple ratees, varying appraisal purposes (however, the motivational consequences of the different purposes have seldom been present), time pressure or constraints, supervisors having performed the major tasks of the subordinate's job, countable results not being available, and raters feeling confident in their ratings (although not discussed, there has typically been high rater confidence in reported lab studies). However, as nice as it is to have a fair amount of overlap between study and modal characteristics, there are significant modal criteria which have not been included in much of the cognitive research. A number of the neglected modal criteria relate to rater-ratee interactions and resulting motivational consequences. Very few, if any, studies in the cognitive area have had ratees sign off on the rating form and indicate their agreement or disagreement, ratees who are aware of how fellow workers are evaluated, or supervisors who are responsible for subordinate's work products. In addition, most studies do not include organization policy variables, such as conducting ratings on an annual basis, having a supervisor's ratings signed off at the next highest level, or having raters evaluating ratee traits. Future research should take great effort to include this subset of modal characteristics which have been neglected so far.

A final point regarding the generalizability of the present studies relates to the emphasis placed on subjective organization strategies. This issue has not received as much attention as the other issues, but may be important to consider now that organization has been identified as a key variable in rating accuracy. What needs to be addressed is whether or not the categorization and organizational processes identified here are indeed similar to those actually used by raters in organizational settings. There may be more potential organizing cues available in the field than investigated in the laboratory. For instance, when making a rating for tenure, there may be only two categories used: tenurable or not. Likewise, there may be categories, such as "good citizen" that are relevant for specific purposes, but not for others. Thus, categories may vary substantially as a function of appraisal purpose. The categories studied to date have come from social cognition research and hence from the laboratory. There may be different types of categories used in the field. The next step for future research to take should be to identify some of the categories actually used in the field. Thus, it may be unrealistic to simply



transport laboratory methods and interventions intact into the field and expect substantial improvements in rating accuracy. Field experimentation may be needed first to identify the congruency between information processing in the field and the laboratory.

In conclusion, however, it would seem that cognitive approaches to studying performance appraisals have led us to consider variables that otherwise would have been overlooked. Specifically, cognitive approaches have resulted in researchers studying the subjective organization of information in memory, the distinction between on-line and memory-based ratings, the role of rater schemata, and the problems involved in a rater's reprocessing of information collected for one purpose so that it can be used to make ratings. These variables and processes have provided new insights into the appraisal process and have generated somewhat novel suggestions for interventions aimed at improving ratings. Although there is still a great deal of work to be done, it seems clear that cognitive approaches to studying appraisals have gone far in enabling researchers to better understand how appraisal decisions are made, and therefore how to improve the quality and accuracy of these decisions.

## REFERENCES

- Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, 93, 203-231.
- Anderson, N. H., & Hubert, S. (1963). Effects of concomitant verbal recall on order effects in personality impression formation. *Journal of Verbal Learning and Verbal Behavior*, 2, 379-391.
- Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecalled information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior*, 17, 1-12.
- Balzer, W. K. (1986). Biases in the recording of performance-related information: The effects of initial impression and centrality of the appraisal task. *Organizational Behavior and Human Decision Processes*, 37, 329-347.
- Banks, C. G., & Murphy, K. R. (1985). Toward narrowing the research-practice gap in performance appraisal. *Personnel Psychology*, 38, 335-344.
- Banks, C. G., & Roberson, L. (1983 August). *Relationship between cognitive schema and rating accuracy*. Paper presented at the annual meeting of the American Psychological Association, Anaheim, CA.
- Bargh, J. A. (1982). Attention and automaticity in the processing of self-relevant information. *Journal of Personality and Social Psychology*, 43, 425-436.
- Bargh, J. A. (1984). Automatic and conscious processing of social information. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 2, pp. 1-43). Hillsdale, NJ: Erlbaum.
- Barnes-Farrell, J. L., & Couture, K. A. (1983, August). Effects of appraisal salience on immediate and memory-based judgments. In C. Banks & L. Roberson (Chairs), *Cognitive processes in performance appraisal: New findings*. Symposium presented at the 91st meeting of the American Psychological Association, Anaheim, CA.
- Bartlett, F. C. (1932). *Remembering*. New York: Macmillan.
- Bernardin, H. J. (1979). Rater training: A critique and reconceptualization. *Proceedings of the 39th Annual Meeting of the Academy of Management*, 216-220.
- Bernardin, H. J., & Beatty, R. W. (1984). *Performance appraisal: Assessing human behavior at work*. Boston: Kent.
- Bernardin, H. J., & Buckley, M. R. (1981). A consideration of strategies in rater training. *Academy of Management Review*, 6, 205-212.

- Bernardin, H. J., & Pence, E. C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. *Journal of Applied Psychology*, 65, 60-66.
- Bernardin, H. J., & Villanova, P. (1986). Performance appraisal. In E. A. Locke (Ed.), *Generalizing from the laboratory to field settings* (pp. 43-62). Lexington, MA: Lexington.
- Bernardin, H. J., & Walter, C. S. (1977). Effects of rater training and diary keeping on psychometric errors in rating. *Journal of Applied Psychology*, 62, 64-69.
- Blencoe, A. G. (1984). *An examination of rater's cognitive processes in performance appraisal: A laboratory investigation*. Unpublished Ph.D. dissertation, University of South Carolina.
- Blumberg, H. H., DeSoto, C. B., & Kueth, J. L. (1966). Evaluations of rating scale formats. *Personnel Psychology*, 19, 243-259.
- Borman, W. C. (1978). Exploring the upper limits of reliability and validity in job performance ratings. *Journal of Applied Psychology*, 63, 135-144.
- Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 49, 229-240.
- Bower, G. H., (1981). Mood and memory. *American Psychologist*, 129-148.
- Brown, E. M. (1968). Influence of training, method, and relationship on the halo effect. *Journal of Applied Psychology*, 52, 195-199.
- Cafferty, T. P., DeNisi, A. S., & Williams, K. J., (1986). Search and retrieval patterns for performance information: Effects on evaluations of multiple targets. *Journal of Personality and Social Psychology*, 50, 676-683.
- Cardy, R. L., & Dobbins, G. H. (1986). Affect and appraisal accuracy: Liking as an integral dimension in evaluating performance. *Journal of Applied Psychology*, 71, 672-678.
- Cardy, R. L., & Kehoe, J. F. (1984). Rater selective attention ability and appraisal effectiveness: The effect of cognitive style on the accuracy of differentiation among ratees. *Journal of Applied Psychology*, 69, 589-594.
- Cofer, C. N., Bruce, D. R., & Reicher, G. M. (1966). Clustering in free-recall as a function of certain methodological variations. *Journal of Experimental Psychology*, 71, 858-866.
- Cooper, W. H. (1981a). Conceptual similarity as a source of illusory halo in performance ratings. *Journal of Applied Psychology*, 66, 302-307.
- Cooper, W. H. (1981b). Ubiquitous halo. *Psychological Bulletin*, 90, 218-244.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.
- DeCotiis, T. A., & Pettit, A. (1978). The performance appraisal process: A model and some testable hypotheses. *Academy of Management Review*, 21, 635-646.
- DeNisi, A. S. (1977). *Sampling strategies and several other factors influencing the accuracy of performance evaluations*. Unpublished Ph.D. dissertation, Department of Psychological Sciences Purdue University.
- DeNisi, A. S., Cafferty, T. P., & Meglino, B. M. (1984). A cognitive model of the performance appraisal process: A model and research propositions. *Organizational Behavior and Human Performance*, 33, 360-396.
- DeNisi, A. S., Cafferty, T. P., Williams, K. J., Blencoe, A. G., & Meglino, B. M. (1983). Rater information acquisition strategies: Two preliminary experiments. *Proceedings of the Academy of Management*, 169-172.
- DeNisi, A. S., Gardner, T., & Cafferty, T. P. (1987). Diary keeping and the organization of information for performance appraisals. *Proceedings of the Academy of Management Meetings*.
- DeNisi, A. S., & Summers, T. (1986, August). *Rating forms and the organization of information: A role for appraisal instruments*. Paper presented at annual meeting of the Academy of Management, Chicago.
- DeNisi, A. S., & Williams, K. J. (1986, April). The purpose of performance evaluations: A cognitive interpretation. In J. Cleveland (Chair), *Purpose of performance evaluation: Rater intentions versus organizational uses*. Symposium presented at the annual meeting of the Society for Industrial and Organizational Psychology, Chicago.



- Dreben, E. K., Fiske, S. T., & Hastie R. (1979). The independence of evaluative and item information: Impression and recall order effects in behavior-based impression formation. *Journal of Personality and Social Psychology*, 37, 1758-1768.
- Dunham, R. B. (1983). Organizational practices. *The Industrial Organizational Psychologist*, 21, 46.
- Ebbeson, E. B., & Allen, R. B. (1979). Cognitive processes in implicit personality trait inferences. *Journal of Personality and Social Psychology*, 37, 471-488.
- Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology*, 66, 127-148.
- Feldman, J. M. (1986). Instrumentation and training for performance appraisal: A perceptual-cognitive viewpoint. In K. M. Rowland & G. R. Ferris (Eds.), *Research in personnel and human resource management* (Vol. 4, pp. 45-99). Greenwich, CT: JAI Press.
- Feldman, J. M., Camburn, A., & Gatti G. (1986). Shared distinctiveness as a source of illusory correlation in performance appraisal. *Organizational Behavior and Human Decision Processes*, 37, 34-59.
- Fisher, C. D., (1979). Transmission of positive and negative feedback to subordinates: A laboratory investigation. *Journal of Applied Psychology*, 64, 533-540.
- Gutman, J. (1982). A means-end chain model based on consumer categorization processes. *Journal of Marketing*, 60-72.
- Hamilton, D. L., Katz, L. B., & Leirer, V. O. (1980). Cognitive processes in first impression formation. *Journal of Personality and Social Psychology*, 39, 1050-1063.
- Hansen, R. D. (1980). Commonsense attribution. *Journal of Personality and Social Psychology*, 39, 969-1009.
- Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review*, 93, 256-268.
- Hastie, R., Park, B., & Weber, R. (1984). Social memory. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 2, pp. 151-212). Hillsdale, NJ: Erlbaum.
- Hernstein, J. A., Carroll, J. S., & Hayes, J. R. (1981). The organization of knowledge about people and their attributes in long-term memory. *Representative Research in Social Psychology*, 11, 17-37.
- Hoffman, C., Mischel, W., & Mazze, K. (1981). The role of purpose in the organization of information about behavior: Trait-based versus goal-based categories in person cognition. *Journal of Personality and Social Psychology*, 40, 211-225.
- Ilgen, D. R., & Faverio, J. L. (1985). Limits in generalization from psychological research to performance appraisal process. *Academy of Management Review*, 10, 311-321.
- Ilgen, D. R., & Feldman, J. M. (1983). Performance appraisal: A process focus. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior* (Vol. 5, pp. 141-197). Greenwich, CT: JAI Press.
- Jacoby, J., Chestnut, R. W., Weigel, K. C., & Fischer, W. A. (1976). Pre-purchase information acquisition: Description of a process methodology, research paradigm, and pilot investigation. In B. B. Anderson (Ed.), *Advances in consumer research* (Vol. 3), *Proceeding of the Association for Consumer Research Conference*.
- Johnson, D. M. (1963). Reanalysis of experimental halo effects. *Journal of Applied Psychology*, 47, 46-47.
- Johnson, D. M., & Vidulich, R. N. (1956). Experimental manipulation of the halo effect. *Journal of Applied Psychology*, 40, 130-134.
- Jolly, J. P., Reynolds, T. J., & Slocum, J. W. (in press). Application of the means-end theoretic for understanding the cognitive bases of performance appraisal. *Organizational Behavior and Human Decision Processes*.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.

- Kane, J. S., & Lawler, E. E., III. (1979). Performance appraisal effectiveness: Its assessment and determinants. In B. M. Staw (Ed.), *Research in organizational behavior* (Vol. 1, pp. 425-478). Greenwich, CT: JAI Press.
- Kelley, H. H. (1971). *Attribution in social interaction*. Morristown, NJ: General Learning Press.
- Kinicki, A., & Lockwood, C. A. (1985). *The impact of appraisal purpose on rater accuracy and the encoding of performance information*. Unpublished manuscript, Department of Management, Arizona State University.
- Kozlowski, S. W. J., Kirsch, M. P., & Chao, G. T. (1986). Job knowledge, rater familiarity, conceptual similarity, and halo error: An exploration. *Journal of Applied Psychology*, 71, 45-49.
- Landy, F. S., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, 87, 72-107.
- Landy, F. S., Zedeck, S., & Cleveland J. (1983). *Performance measurement and theory*. Hillsdale, NJ: Erlbaum.
- Lingle, J. H., & Ostrom, T. M. (1979). Retrieval selectivity in memory-based judgments. *Journal of Personality and Social Psychology*, 37, 180-194.
- Loftus, G. R., & Loftus, E. F. (1976). *Human memory: The processing of information*. Hillsdale, NJ: Erlbaum.
- Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology*, 12, 523-553.
- Lord, R. G. (1985). An information processing approach to social perceptions, leadership and behavioral measurement in organizations. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior* (Vol. 7, pp. 87-128). Greenwich, CT: JAI Press.
- Lord, R. G., & Alliger, G. M. (1985). A comparison of four information processing models of leadership and social perceptions. *Human Relations*, 38, 47-65.
- Lord, R. G., Foti, R. J., & Phillips, J. S. (1982). A theory of leadership categorization. In J. Hunt, U. Sekaran, & C. Schriesheim (Eds.), *Leadership: Beyond establishment views*. Carbondale: Southern Illinois University Press.
- Lord, R. G., Foti, R. J., & DeVader, C. (1984). A test of leadership categories, information processing, and leadership perceptions. *Organizational Behavior and Human Performance*, 34, 343-378.
- Matte, W. E. (1981). *An experimental investigation of information search in performance appraisal*. Unpublished Ph.D. Dissertation, Department of Psychological Sciences, Purdue University.
- Mandler, G. (1967). Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 1). New York: Academic Press.
- McArthur, L. Z. (1972). The how and what of why: Some determinants and consequences of causal attribution. *Journal of Personality and Social Psychology*, 22, 171-193.
- McCann, D., Mitchell, M. L., Ostrom, T. M., & Herstein, J. (1983). *Memory for social information: Using persons as organizing categories*. Paper presented at the meeting of the American Psychological Association, Washington, D. C.
- McCauley, C., & Stitt, C. C. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology*, 36, 929-940.
- Meyer, H. H., Kay, E., & French, J. (1965). Split roles in performance appraisal. *Harvard Business Review*, 43, 123-129.
- Mosier, C. I. (1940). Psychophysical and mental test theory: Fundamental postulates and elementary theorems. *Psychological Review*, 47, 355-366.
- Murphy, K. R., & Balzer, W. K., (1986). Systematic distortions in memory-based behavior rating and performance evaluations: Consequences for rating accuracy. *Journal of Applied Psychology*, 71, 39-44.
- Murphy, K. R., Balzer, W. K., Lockhart, M. C., & Eisenman, E. J. (1985). Effects of previous performance on evaluations of present performance. *Journal of Applied Psychology*, 70, 72-84.
- Murphy, K. R., Gannett, B. A., Herr, B. M., & Chen, J. A. (1986). Effects of subsequent performance on evaluations of previous performance. *Journal of Applied Psychology*, 71, 427-431.

- Murphy, K. R., Garcia, M., Kerkar, S., Martin, C., & Balzer, W. K. (1982). Relationship between observational accuracy and accuracy in evaluating performance. *Journal of Applied Psychology*, 67, 320-325.
- Murphy, K. R., Herr, B. M., Lockhart, M. C., & Maguire, E. (1986). Evaluating the performance of paper people. *Journal of Applied Psychology*, 71, 654-661.
- Murphy, K. R., Martin, C., & Garcia, M. (1982). Do behavior observation scales measure observation? *Journal of Applied Psychology*, 67, 562-567.
- Nisbett, R. E., & Borgida, E. (1975). Attribution and the psychology of prediction. *Journal of Personality and Social Psychology*, 32, 932-943.
- Nathan, B. R., & Lord, R. G. (1983). Cognitive categorization and dimensional shemata: A process approach to the study of halo in performance rating. *Journal of Applied Psychology*, 68, 102-114.
- Newton, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, 28, 28-38.
- Newton, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, 12, 436-450.
- Newton, D., Engquist, G., & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology*, 35, 847-862.
- Ostroff, C., & Ilgen, D. R. (1986, April). *The relationship between cognitive categories of raters and rating accuracy*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Chicago.
- Ostrom, T. M., Pryor, J. B., & Simpson, D. D. (1981). The organization of social information. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario symposium on personality and social psychology* (pp. 3-38). Hillsdale, NJ: Erlbaum.
- Posner, M. I., & Snyder, C. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium*. Hillsdale, NJ: Erlbaum.
- Pryor, J. B., & Ostrom, T. M. (1981). The cognitive organization of social information: A converging-operations approach. *Journal of Personality and Social Psychology*, 41, 628-641.
- Pryor, J. B., Ostrom, T. M., Dukerich, J. M., Mitchell, M. L., & Herstein, J. A. (1983). Preintegrative categorization of social information: The role of persons as organizing categories. *Journal of Personality and Social Psychology*, 44, 923-932.
- Pulakos, E. D. (1984). A comparison of two rater training programs: Error training versus accuracy training. *Journal of Applied Psychology*, 69, 581-588.
- Pulakos, E. D. (1986). The development of training programs to increase accuracy with different rating tasks. *Organizational Behavior and Human Decision Processes*, 38, 76-91.
- Pulakos, E. D., Schmitt, N., & Ostroff, C. (1986). A warning about the use of a standard deviation across dimensions within rates to measure halo. *Journal of Applied Psychology*, 71, 219-32.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104, 192-233.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1-66.
- Schul, Y. (1983). Integration and abstraction in impression formation. *Journal of Personality and Social Psychology*, 44, 45-54.
- Sharon, A. T., & Bartlett, C. J. (1969). Effect of instructional conditions in producing leniency on two types of rating scales. *Personnel Psychology*, 22, 252-263.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127-190.
- Srull, T. K. (1983). Organizational and retrieval processes in person memory: An examination of processing objectives, presentation format, and the possible role of self-generated retrieval cues. *Journal of Personality and Social Psychology*, 44, 1157-1170.
- Srull, T. K., & Brand, J. F. (1983). Memory for information about persons: The effect of encoding operations on subsequent retrieval. *Journal of Verbal Learning and Verbal Behavior*, 22, 219-230.

- Srull, T. K., & Wyer, R. S. (1980). Category accessibility and social perception: Some implications for the study of person memory and interpersonal judgment. *Journal of Personality and Social Psychology*, 38, 841-856.
- Taylor, E. K., & Hastman, D. (1956). Relation of format and administration to the characteristics of graphic rating scales. *Personnel Psychology*, 9, 181-206.
- Thornton, G., & Zorich, S. (1980). Training to improve observer accuracy. *Journal of Applied Psychology*, 65, 351-354.
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4, 25-29.
- Tiffin, J., & McCormick, E. J. (1965). *Industrial psychology* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Tulving, E. (1974). Recall and recognition of semantically encoded words. *Journal of Experimental Psychology*, 102, 778-787.
- Weiner, B., Frieze, I., Kukla, A., Reed L., Rest, S., & Rosenbaum, R. M. (1971). Perceiving the causes of success and failure. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, NJ: General Learning Press.
- Wherry, R. J. (1952). *The control of bias in ratings: A theory of rating*. Columbus: The Ohio State Research Foundation.
- Wherry, R. J., & Bartlett, C. J. (1982). The control of bias in ratings: A theory of rating. *Personnel Psychology*, 35, 521-551.
- Williams, K. J. (1984). *Appraisal salience and the need to reprocess information: Effects on organization, recall, and rating of performance information*. Unpublished Ph.D. dissertation, Department of Psychology, University of South Carolina.
- Williams, K. J. (1987). *The effects of structured and unstructured diaries on performance ratings*. Unpublished manuscript, Department of Psychology, SUNY-Albany.
- Williams, K. J., Cafferty, T. P., & DeNisi, A. S. (1985). Appraisal salience: I. The effects of reprocessing objectives on recall and rating of performance information. *Proceedings of the Southern Management Association Meetings*, 121-123.
- Williams, K. J., DeNisi, A. S., Blencoe, A. G., & Cafferty, T. P. (1985). The role of appraisal purpose: Effects of purpose on information acquisition and utilization. *Organizational Behavior and Human Decision Processes*, 36, 314-339.
- Williams, K. J., DeNisi, A. S., Meglino, B. M., & Cafferty, T. P. (1986). Initial decisions and subsequent performance ratings. *Journal of Applied Psychology*, 71, 189-195.
- Williams, K. J., & Keating, C. W. (1987, April). Affect and the processing of performance information. In G. H. Dobbins (Chair), *Affect in human resource management: Implications for appraisals, promotions, and sanctions*. Symposium presented at the annual meeting of the Society of Industrial and Organizational Psychology, Atlanta.
- Williams, K. J., Wickert, P., & Peters, R. D. (1985). Appraisal salience: II. Effects of instructions to subjectively organize information. *Proceedings of the Southern Management Association Meetings*, 124-126.
- Wyer, R. S., & Srull, T. K. (1981). Category accessibility: Some theoretical and empirical issues concerning the processing of social stimulus information. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario symposium* (pp. 161-197). Hillsdale, NJ: Erlbaum.
- Wyer, R. S., Srull, T. K., & Gordon, S. (1984). The effects of predicting a person's behavior on subsequent trait judgments. *Journal of Experimental Social Psychology*, 20, 29-46.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35, 151-175.
- Zedeck, S., & Cascio, W. (1982). Performance appraisal decisions as a function of rater training and purpose of the appraisal. *Journal of Applied Psychology*, 67, 752-758.