Beyond Attribution Theory:
Cognitive Processes in Performance Appraisal

Jack M. Feldman
Department of Management, University of Florida

Performance appraisal is construed as the outcome of a dual-process system of evaluation and decision making. Attention, categorization, recall, and information integration are carried out via either an automatic or a controlled process. In the automatic process, aspects of an employee and his/her behavior are noted and the employee is categorized without conscious monitoring. The automatic process is dominant except when decisions are problematic, in which case a consciously monitored categorization process takes place. Subsequent recall of the employee is biased by the attributes of prototypes (abstract images) representing categories to which the employee has been assigned. Dispositional and contextual factors influence the availability of categories during both assignment and recall. Categorization also biases any subsequent search for information about the employee, and interacts with task type to produce halo, leniency/stringency biases, and racial, sexual, ethnic, and personalistic bias as well. The same automatic and controlled processes can, however, account for accuracy of evaluations. Suggested research includes the study of behavior taxonomies, individual differences in cognitive structure, the validation of behavior-sampling techniques, and laboratory studies of appraisal processes.

In any organization, some of the most important decisions concern people: Who is the right person for the job? Who should be promoted? Who deserves a raise? Who can take on more responsibility? These questions are sometimes addressed formally, by using various types of evaluation procedures (see, e.g., Campbell, Dunnette, Lawler, & Weick, 1970; Smith, 1976). More often, these decisions are made informally, even though they have a great impact on the employee's career (Dansereau, Graen, & Haga, 1975; Graen, 1976).

There are two separate bodies of research bearing on the issue of employee evaluation, one concerned with the instrument and the other with its user. The first, in the classic psychometric tradition, centers on the development of formal evaluation instruments and their susceptibility to both random error and various systematic biases (such as halo effects, leniency/stringency, central tendency, and the like). The other area of research is based in social psychology and is concerned with attribution and stereotyping processes and their influence on evaluation (especially racial, cultural, and sexual biases; see Bigoness, 1976; Feldman & Hilterman, 1977; Hamner, Kim, Baird, & Bigoness, 1974; Rosen & Jerdee, 1974; Terborg & Ilgen, 1975). Recently, Stone and Slusher (1975) have tried to link the psychometric and social areas by describing performance appraisal techniques in attributional terms. Green and Mitchell (1979) and Bazerman and Atkin (Note 1) have applied attribution theory directly to the performance appraisal process.

The purpose of this article is to suggest a new approach to perfomance appraisal, one with implications for both the psycho-

metic and social-psychological traditions. A "model of the rater" (Borman, 1978, pp. 141–143; Atkin & Conlon, Note 2) is proposed that will provide a single conceptual tool for analyzing several distinct problems: What cognitive processes are engendered by various types of rating scales now in use? How are informal, everyday evaluations of and decisions about employees made? How can a single model explain the fact that judgments contain both valid information about individual differences in job behavior yet also contain halo errors, leniency/stringency/central tendency biases, and possible bias due to race, sex, ethnicity, and other personal attributes of employees?

Answers to these questions should serve to guide future research on the appraisal process, the development of new evaluation forms and systems, and the selection and training of evaluators. It is assumed that performance appraisal is a specific case of more general cognitive processes and that to fully understand it, these general processes must be considered.

The Appraisal Context

A brief description of the context in which performance appraisals (formal and informal) are conducted may help organize the issues. The typical manager has many duties, only some of which involve the direct supervision of subordinates (e.g., Dowell & Wexley, 1978; Mintzberg, 1973). Direct information about subordinates' job behavior is often fragmentary; direct personal contact with subordinates may be minimal and restricted to a particular set of situations, depending on the nature of the job. Jobs themselves are incompletely understood, and specific duties may be inadequately described or entirely unspecified, especially at higher managerial levels (Campbell et al., 1970). In this uncertain, informationally "noisy" environment, the supervisor must perform several cognitive tasks before performance appraisals are possible:

1. The supervisor must recognize and attend to relevant information about employees.

2. This information must be organized and stored for later access. New information must also be integrated with previously gathered data.

3. When judgments are required—for example, when work assignments are to be made—relevant information must be recalled in an organized fashion.

4. At various times in the above stages, or when task demands (such as evaluation forms) require it, information must be integrated into some sort of summary judgment.

It should be evident that performance appraisal, as envisioned here, is a complex and cyclical process or rather a combination of interacting processes. Time is a vital consideration, since impressions and evaluations are formed as behavior is observed sequentially, and final judgments are based as much or more on memory as on current observation.

Previous studies of attributional processes in performance evaluation (cited above) are deficient in that they deal exclusively with the integration of information that is complete and immediately available. Although relevant, a focus on such stimulus-based judgments neglects the selection of information from the environment and the organization, storage, and recall processes necessary to more typical memory-based judgments. These studies also do not address a vital issue: the circumstances under which conscious, deliberate attributions do not occur at all (Enzle & Schopflocher, 1978). To handle these issues, this article discusses automatic and controlled (consciously monitored) processes of attention, search, stimulus detection, and perceptual learning, and their analogs in social behavior; the nature of categories and categorization processes, both automatic and controlled; category accessibility and the influence of categorization on subsequent judgments; and the implication of these topics for research and practice in performance appraisal.

The article is organized around the four cognitive tasks described above, in an order roughly approximating the order in which appraisals seem to be formed. This ordering is necessarily imperfect, due to the cyclical and interactive nature of the overall conceptual model and the fact that both automatic and controlled processes must be discussed.

Attention and Recognition Processes

Schneider and Shiffrin (1977) and Shiffrin and Schneider (1977) recently made a strong case for the presence of automatic and controlled processes in attention, search, detection, and perceptual learning, all of which are relevant to performance appraisal. To oversimplify, under conditions in which one is required to recognize or detect a previously seen stimulus from among a set of distracting stimuli, either an automatic or controlled process takes place. The automatic process occurs under conditions of consistent mapping, in which a given stimulus type (say, numbers or letters) is either always a target or always a distractor, but not both. The person is set to note or attend to particular stimulus configurations, and this takes place without continuous monitoring. Under conditions in which a given stimulus type may be either a distractor or a target (variable mapping), a controlled process is evoked that requires conscious direction and decision making.

Recent developments in the literature of cognition and attribution make plausible the notion that similar automatic and controlled processes take place in social interaction. Langer (1978) and Nisbett and Wilson (1977) both show that everyday behavior may take place in a thoughtless or mindless fashion, suggesting that people respond to cues of which they are unaware. As Gibbs (1979) points out, this does not mean that people are not thinking, only that they are not thinking about what they are doing at the time. The notion of "script processing" (Abelson, 1976; Shank & Abelson, 1977) is consistent with this view, since the performance of well-learned sequences of behavior provides ample opportunity for a person's attention to be directed away from the behavior of the moment.

The implication of this automaticity is straightforward. People learn to attend to certain stimulus features without monitoring this process. Race, sex, cues of dress and speech, height, and so on, are all stimuli that can be automatically recorded. People may be categorized via such stimuli (as discussed below) without intention but with future consequences for the interpretation of their behavior. That is, one does not ask, "Is that a woman, and what difference does it make if she is?". One recognizes femaleness automatically and unintentionally and thereafter reacts partially in terms of that classification.

Taylor and Fiske (1978) further note that people's impressions of themselves and others are unwittingly influenced by factors that make certain kinds of information (e.g., race or sex) more salient or prominent in the perceptual field (see also McArthur & Post, 1977). As Langer (1978) points out, it is only when an interaction becomes problematic or effortful that it requires attention; likewise, it may only be when an inference about other people becomes problematic or effortful that the controlled attention and search processes are brought into play (see Newtson, 1976). In other words, the person may use the automatic process until a change from consistent to variable mapping conditions becomes noticeable, perhaps because mistakes of classification are being made (as when someone who looks like a good employee turns out to be a poor one). Even after the controlled process is invoked, the data recalled and brought to bear on the problem may have been influenced by the previous automatic processes.

In summary, to the degree the behavior of an employee is consistent with the supervisor's expectations, it is noted and stored automatically. It is only when a behavior departs from expectations, or when the task is somehow changed, that conscious attention and recognition processes are engaged.

It is true that behavior, as a stimulus, differs greatly from the stimuli employed in studies such as Schneider and Shiffrin's (1977). For one thing, any single behavior may be ambiguous, having several possible interpretations. How behavior is interpreted and stored and when automatic or controlled processes are engaged are discussed below.

Information Organization and Storage

The Nature of Categories and Categorization

Categorization is basic to perception (Bruner, 1958), information storage, and organization. It enables people to process large amounts of data despite inherent lim-

itations of storage capacity (see Smith, Adams, & Schorr, 1978).

Rosch and her coworkers (Rosch, 1977; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) hold that categories are not (as previous research assumed) firmly bounded by singly necessary and jointly sufficient sets of attributes. Rather, categories are defined as "fuzzy sets." A category member's possession of attributes is probabilistically related to category membership; the possession of every relevant attribute is unnecessary for membership. Rosch et al. propose that categorization of a stimulus is based on the extent to which the features of the given stimulus overlap with those of a category prototype (Tversky, 1977), an abstract analog or image summarizing family resemblances among category members. In this sense, category exemplars may be said to be more or less prototypical. For example, a plumber is closer to the prototypical skilled worker than is a paramedic or a laboratory technician. Categories develop along the lines of the correlations found in nature in which, for example, overalls are more typically associated with hand tools than is a white coat. Thus, the prototypical skilled worker will be seen as wearing overalls and using hand tools.[1]

Prototypes and Inferences About People

Once a stimulus is categorized, recall and recognition of that stimulus are biased toward the general characteristics of the category, including the recognition of information that was never in fact presented (Cantor & Mischel, 1977, 1979; Tsujimoto, 1978; Tsujimoto, Wilde, & Robertson, 1978). The assignment of a person to a category under consistent mapping conditions, where little ambiguity exists, may be considered an automatic process. Although memory search itself is a controlled process, the above-mentioned memory biases are due to differential forgetting and information accessibility; as such, they too may be considered automatic in the sense that no decision is required for their operation. However, Srull (Note 4) has raised the possibility that when specific information is unavailable, the prototype is used as a basis for guessing. Thus, a person-

memory task may have problem-solving components and therefore qualify as a controlled rather than an automatic process. Whether people mistakenly remember, or infer and report remembering, unseen information is an important question that cannot be resolved at present. Either process would be consistent with Wyer and Srull's (1980) memory and information processing model (see below).

The relevance of this process to performance evaluation is also simple. When an employee is assigned to a category, further memory-based judgments of that employee are colored by the category prototype. This process, functionally identical to stereotyping (Ashmore & DelBoca, 1979) can produce either underevaluations or overevaluations of employees by associating the general evaluation of the category with the person, producing false memories of the person, or both. This would be especially true where a supervisor has a large number of employees to evaluate, little chance for extended interaction with them, and the evaluation form is a simple evaluative rating, a trait rating, or a recognition task such as a behaviorally anchored rating scale or behavioral checklist. Note that neither overt prejudice nor motivational biases are necessary to explain such results, which arise from the nature of the categorization process itself.

Thus, we can say that under conditions allowing the immediate assimilation of a given person to a category prototype, further information about that individual will be automatically interpreted and stored in terms of that same category, which becomes more accessible than others (Wyer & Srull, 1980).

Category selection: Situational factors. Having outlined the simplest case of the categorization process (in which a person fits a single category based on race, sex, occupation, etc.), complications must now be introduced. It is obvious that people are quite complex stimuli and that any given person may be categorized in several ways. Further,

---

[1] For further discussion of categorization processes, see McCloskey and Glucksberg (1979), Medin and Schaffer (1978), and Smith and Medin (Note 3).

characteristics of the perceiver are also important to the categorization process.

Looking at the stimulus side first, we may ask how any given aspect of a stimulus person comes to be used as a basis for categorization. Is a soft-spoken, black, college graduate, age 22, categorized as an introvert, a black, a graduate, or a young person? Taylor and Fiske (1978) have shown convincingly that aspects of the situation that make a given feature of a stimulus person more salient, in the perceptual sense, strongly influence this process. The only female or black in a group is seen as more typical of females or blacks and more causal in group affairs than that same individual in the company of other blacks, females, and so on. Other salience manipulations (seating position, the presence or placement of cameras, and frequency or novelty of stimuli) may produce similar effects. One may postulate that situational factors influence the salience of given categories and their prototypes, directing attention to prototype-relevant stimulus features and away from others.

Similarly, Tversky (1977) has shown that changing judgment contexts can dramatically change the perceived similarity of two objects by changing the relative degree of feature overlap between any two objects in a set. This, too, can be seen as an example of category salience, in that features possessed in common by a set of stimuli may evoke one category prototype rather than another. This is an important issue in performance appraisal, for any model must deal with the evaluation of all employees, not simply those who immediately and obviously belong to only one particular category. Consider the token; whether the person is the only black in an all-white group, the only male in an all-female group, the only tieless person in a room of three-piece suits, and so on, the very fact of difference focuses attention on both the person and the dimension of difference. This makes certain categories more likely to be used and particular evaluations or attributions more likely than others.

Category selection: Individual differences. On the perceiver's side, Markus (1977) has shown that schemata (prototypes) about the self organize self-relevant information and that people vary in the schemata they possess. Just so, people vary in

terms of the number and nature of the other categories they possess. These categories and their associated prototypes are called either stereotypes or implicit personality theories depending on whether they are based on cues such as race or on trait labels such as extrovert (see Hastorf, Schneider, & Polefka, 1970).

Kelly's (1955) personal construct theory delineated the sometimes profound individual differences that exist in category systems. Rosenberg and Sedlak (1972) have shown these category systems to be representable by hierarchial clustering models similar to those suggested for other types of category systems (Rosch, 1977). Further, it is thought (Tajfel, 1969) that the affect associated with a category is one determinant of its salience. It is likely that, due to cultural factors (Triandis, 1964) and individual difference variables such as prejudice and cognitive complexity (Feldman & Hilterman, 1975), different categories will be salient for different people. These individual differences will, in conjunction with situational factors (such as how recently a category has been used; Wyer & Srull, 1980), determine which aspects of a given stimulus person will be used in categorization, other things being equal.

The relevance of these individual differences to performance appraisal is straightforward. To the extent that information about employees is stored and recalled in terms of category prototypes, the actual category system used by an evaluator will play a large role in determining subsequent evaluations. Since categories are formed in terms of covariations observed by the individual as well as by direct tuition, there will be both unique and culture-common variance in each person's system. This can cause problems when supervisors and workers are from different cultural backgrounds. For example, if the employee's "friendly" prototype contains aspects of the supervisor's "disrespectful" one, the employee's attempts to show friendship may have harmful (or possibly amusing) consequences.

Controlled Categorization Processes

What if other things are not equal? So far the automatic categorization process, en-

gaged when stimuli fit consistently into a given category, has been discussed. What about the more problematic situations that cannot be handled thoughtlessly (i.e., where there is no immediately available prototype to which the person can be satisfactorily matched)? This is the point at which controlled categorization processes must be introduced.

A controlled categorization process is triggered when incoming information reaches some hypothetical threshold of discrepancy. That is, when the observed behavior of the individual in question departs sufficiently from that expected on the basis of an initial categorization, a problem-solving recategorization process must be brought into play (Feldman, 1972; Feldman & Hilterman, 1975; Jones & McGillis, 1976). Likewise, when the person being observed does not easily match any available prototype, more information must be gathered and a decision made.

This hypothetical triggering is consistent with Newtson's (1976) data showing that smaller units of behavior are attended to by the perceiver when attributions are uncertain, Langer's (1978) finding that people in everyday situations do not respond to the content of communication until interaction becomes effortful or problematic, and Hastie's (1978) demonstration of greater recall for category-incongruent than category-irrelevant behavior. Such processes have usually been discussed under the label of "information integration" (Anderson, 1974), in that different pieces of information must be combined to yield a summary judgment. Attribution theory, which concentrates on stimulus-based judgments, seems well suited to describe the process that occurs when categorization cannot proceed automatically. It will also become relevant at a later point, when formal appraisals are being considered. Now the issue is one of initial categorization.

As described by Kelley (1971a, 1971b), Jones (1979), Jones and Davis (1965), Jones and McGillis (1976), Weiner (1972), and others, attribution is a rational process in which people behave as naive scientists when drawing conclusions about others. Two principles are fundamental to attribution theory. The first, covariation, states that the perceived cause of an event is found among the conditions varying with the occurence of the event rather than among those that are unchanging (see Shaver, 1975). Kelley's (1971a) "naive analysis of variance" contains three dimensions over which events or behaviors may vary: entities, or things acted toward, such as tasks; situations or contexts of behavior, such as stressful work settings, the golf course, or a training program; and other people—that is, in what ways does the behavior of the person in question depart from the base rate of behavior across situations and entities?

The second principle of attribution, discounting, states that an observer will discount any single causal explanation for a given behavior when more than one plausible cause is available (see Kruglanski, 1970; Strickland, 1958).

The first stage in the controlled process is causal attribution: The perceiver must decide if the event or behavior is caused by the person in question or is due to the situation or entity involved. If the person covaries with the behavior or event, he or she is seen as the cause. For example, according to the covariation principle, an employee who regularly fails to complete an assigned task when other employees complete similar ones will be seen as the cause of the failure. If only a particular kind of task is not completed, however, and many other employees also fail to complete these tasks, the task will be seen as the cause of the failure. Only in the former case would failure be expected to impact negatively on an employee's evaluation.

An example of the discounting principle in causal attribution would be a case in which an employee, previously categorized as a member of a low-ability group, failed at some new task. The failure would be seen as caused by the employee, and task causality would be discounted.

Smith and Miller (1979) provide evidence that the causal attribution process is subtractive; that is, people begin with complex explanations for behavior and eliminate potential causes as the evidence warrants. They

also suggest that the set of potential causes that perceivers begin with is determined by the salience of those causes (as discussed by Taylor & Fiske, 1978) or the cause's presence in a script of the behavior setting (Shank & Abelson, 1977). This supports the idea that the outcome of controlled attribution processes is partially determined by the automatic attention and categorization processes that have operated earlier.

Once causality has been determined, a second judgment is necessary. This may be termed a trait or dispositional attribution, or a category assignment. The specific category assignment made to the (stimulus) person will be a function of the perceiver's implicit personality theory, or category system, and of situational factors rendering some categories more readily available than others (Srull & Wyer, 1979). Cantor and Mischel (1979) have shown that judgments of prototypicality are made in a manner fully consistent with attribution theory, supporting the idea of dispositional attribution as a categorization process.

In addition to the covariation and discounting principles, attribution theory includes a number of ad hoc biases or moderating factors necessary to encompass the observed data. Ross (1977) has enumerated these biases, some of which can be understood as consequences of the automatic attention and categorization processes discussed earlier.

First, the fundamental error is the tendency to underestimate the importance of situational factors and overestimate that of dispositional factors as causes of behavior. That is, people are seen as causes too frequently, and situations are seen as causes too seldom. In the present context, this can be explained by noting that situational prototypes tend to be characterized by features describing the behavior and dispositions of the people in them (Bem & Funder, 1978; Cantor, Note 5). Parties, for example, are characterized by happy people dancing and talking. Thus, a person's mere presence in a situation has a dispositional connotation.

Second, actors and observers differ in their causal attributions: Actors emphasize situational factors, observers, personal dispositions of the actor. In large measure this tendency is due to the differential salience of situational and personal information to the actor and observer, respectively.

The third bias has already been discussed—the tendency to see as causal the most salient features of the environment, including novel individuals.

The fourth is termed "hedonic relevance" by Jones and Davis (1965): the tendency to see actions having affective consequences for the observer as more dispositional than other actions. Coupled with this is the tendency to see people as more responsible for acts with serious as opposed to trivial consequences. This, too, may stem from the consequent salience of the actor.

Fifth, the actor is held more responsible for acts leading to reward than for acts preventing losses. Also noted is the tendency for people to pay little attention to base rate information (the degree to which a behavior is common) and the complementary tendency for people to use their own behavior as a basis for judging the typicality of an actor's behavior (the "false consensus effect"; see Ross, Greene, & House, 1977). This may be another reflection of the "law of small numbers" (Tversky & Kahneman, 1971), connoting a general tendency to end the search for causal explanation at the first prototype providing a plausible explanation for the behavior.

A last bias, not mentioned by Ross, is for attributions to follow affective relationships: good actions by liked people are attributed to the person and bad actions to the situation, whereas attributions for the actions of disliked people follow the reverse pattern (good actions—external attribution, bad actions—internal; Regan, Strauss, & Fazio, 1974). Taylor and Jaggi (1974) have shown that the same pattern exists in the perception of ethnic groups; it may also underly some of the attributions to the physically attractive and unattractive (Berscheid & Walster, 1974; Cash, Gillen, & Burns, 1977) and explanations for the success of women (Garland & Price, 1977). This may depend, at least partially, on the initial categorization of the person (Srull & Wyer, 1979) and the operation of the discounting principle.

There are two situations in which the controlled attribution process is likely to be used. The first is on initial acquaintance with an employee, provided that circumstances do not permit an automatic categorization. The second is when some event exceeds the perceiver's threshold of discrepancy, requiring a recategorization of the employee. Both cases require judgments that are relatively more stimulus based in that the behavior of the employee serves as a major input to the final categorization. An example of the former case would be a trial period for a new employee who is a member of no ethnic, minority, or social group having a prominent place in the supervisor's category system. An example of the second would be a startlingly good performance by a member of some disadvantaged group (e.g., Feldman & Hilterman, 1975). Mitchell, Green, and Wood (1981) present a thorough theoretical discussion and data supporting the validity of the attributional approach in situations calling for stimulus-based judgments.

Thus, both the automatic and controlled processes have the same end result: the assignment of a person to a category based on a prototype-matching process. The difference is in whether the stimulus person's behavior is sufficiently consistent with other cues to allow the categorization to proceed automatically or whether a controlled process must be used to determine which category is consistent with the individual's behavior. The actual category assignment is a function of contextual factors influencing the salience of particular categories and stimulus characteristics, as well as individual differences among perceivers that render some categories and their prototypes more available than others and some stimulus features more salient than others.

Another issue that should be addressed is degree of acquaintance. The primary differences between the perception of well-known persons and more casual acquaintances in this system are hypothesized to be that well-known people may be assigned to a larger number of more specific categories than casual acquaintances, and prototypicality judgments of acquaintances are made on the basis of their possession of specific and highly central category attributes, whereas well-known people are categorized on the basis of the number and configuration of category-relevant attributes displayed relative to the total number of attributes and on the possession of attributes incompatible with particular categories (Cantor & Mischel, 1979). Thus, the longer a person is known and the more opportunities there are to observe him or her, the more complex the judgment process may become.

It should be noted that this latter process may not occur in every case. It is possible that, having achieved a satisfactory initial categorization, the perceiver turns to other concerns. This would be expected, for example, where the initial categorization was evaluatively negative, and no subsequent events stimulated a reevaluation. In such a case, the person would not become well known because interaction would remain superficial and infrequent. Such may be the case with *out* employees, as described by the vertical dyad model (Dansereau et al., 1975; Graen, 1976).

Summary

Person perception, the basis of performance evaluation, is conceptualized as a dual-process system. Based on studies of perception, cognition, and social behavior, perceivers are said to assign persons to categories. These categories are fuzzy sets, defined by family resemblances among their members and exemplified by category prototypes or images. Stimulus persons may be assigned to categories automatically, by virtue of their possession of obvious or salient attributes; the specific categories to which they are assigned is a function of perceiver and situational factors (e.g., personal constructs and contextual salience). This automatic process may be superseded by a controlled, or consciously monitored, process described by attribution theory, under either of two conditions: when no salient category prototype provides a satisfactory fit or when information sufficiently discrepant with initial categorization is obtained.

Categorization affects performance evaluation by limiting and selecting information

about the employee when memory-based judgments are made and by influencing stimulus-based judgments through the operation of attributional bias.

The Organization of Information Seeking and Recall

Recall

As shown above, the prototype-matching process assigns a stimulus person to a particular category, which thereafter influences the recall of information about that person. Wyer and Srull (1980) have developed a model of information processing and memory incorporating this phenomenon. Briefly, behavioral information is thought to be held for a time in a "workspace"; during this time, information about both the behavior and the category to which the person is assigned is available. After a period of time, the workspace is cleared, and only the categorical information is retained, in one or more "storage bins." Thus, when input information is consistent with a prototype, recall is limited to the stimulus person's categorization. When the person is recalled, behavioral specifics are not stored. The prototype of the relevant category is elicited, and features of the prototype are remembered as being true of the person whether or not relevant behavioral information has ever been presented. If the person is recalled as a prototypical salesman, for example, the trait talkative may be associated with him, since this trait is commonly associated with salesmen in general.

Consider a supervisor completing an employee's annual evaluation. The employee has been assigned to one or more categories. At any given time, some categorizations may be more available than others, due to recency of use, contextual factors, dispositions of the supervisor (Wyer & Srull, 1980), or the supervisor's emotional state (Teasdale & Fogarty, 1979). If the supervisor attempts to remember relevant job-related behaviors of the individual, the behaviors likely to be recalled (and seen as characteristic of the person) are those that are consistent with the evoked prototype.

According to Tversky and Kahneman's (1974) "representativeness" heuristic, people judge the probability of event B being generated by some process A to be greater the more A resembles B. This implies that, for example, a careless behavior is seen as more likely to be performed by a person who is characterized by a prototype having careless as one of its attributes. Their "availability" heuristic postulates that the probability or frequency of an event is judged by the ease with which the event can be brought to mind (its salience). Both heuristics strongly imply that a supervisor will judge the future behavior of an employee on the basis of that prototype that is most easily evoked, or most salient, for the person. Both heuristics have obvious and direct implications for performance evaluations based on recall or recognition of behavior, such as behaviorally anchored rating scales or behavioral observation scales. Recalled prototypes, and behaviors associated with them, would also be expected to guide the generation of critical incidents about employees and the writing of free-response, essay-type evaluations.

Information Search

A conscientious supervisor may seek to confirm his or her impressions of the employee by seeking new information, either directly as in a management by objectives (MBO) interview, or indirectly, through observation or discussion with others. It has been shown that when people expect certain behaviors from a stimulus person, they notice and recall those behaviors more than they do unexpected, but equally available, behaviors (Zadny & Gerard, 1974). Further, when a person is asked to test an hypothesis about another, his or her behavior (in terms of the kinds of questions asked) is such as to seek the confirmation of that hypothesis (Einhorn & Hogarth, 1978; Snyder & Cantor, 1979; Snyder & Swann, 1978a). Also, a person holding firm expectations about the behavior of another is likely to elicit that kind of behavior from the other person in interaction (Snyder & Swann, 1978b).

Thus, categorization not only selectively influences and biases recall, it prevents con-

tradictory evidence from appearing and, in fact, elicits confirming evidence. It almost seems that Kelley's (1971a) naive scientist attribution model should be supplemented by a naive lawyer model of information seeking, in that people seem to act as advocates as well as scientists. In short, when one seeks to test an impression, one is essentially seeking to support that impression, making disconfirmation less likely. It should be noted that this tendency is not limited to supervisors but appears wherever formal attempts to disconfirm hypotheses are not used (Meehl, 1978; Tversky & Kahneman, 1971).

Information Seeking and Recategorization

Taken by itself, the above evidence would imply that categorizations rarely or never change. Yet we know this is untrue from our everyday experience. To explain recategorization, attribution theory must again be invoked. It is known (see Ross, 1977) that a single piece of inconsistent evidence tends to be discounted in an overall impression and that this tendency is greater the more discrepant that single bit of evidence is from the rest. Yet behavior inconsistent with a categorization is the trigger postulated to evoke the controlled attribution process in place of the automatic one. At this point, little is known about such a triggering mechanism. It may be that a single extremely discrepant behavior (e.g., the respected banker embezzling funds) or a series of frequent less-discrepant behaviors are sufficient to exceed the threshold of discrepancy proposed earlier. It may also be that people differ in their thresholds, with greater cognitive complexity, for example, perhaps indicating a lower threshold (see Feldman & Hilterman, 1975). The studies cited earlier do strongly hint at the usefulness of a two-process approach (Bell, Wicklund, Manko, & Larkin, 1976; Hastie, 1978; Langer, 1978; Newtson, 1976).

Interestingly, recategorization of a person may cause a reconstruction of memories about that individual such that memories consistent with the new categorization become more accessible (Snyder & Uranowitz, 1978). When information about the person is sought from memory at a later date, available information seems to consistently support the new category placement.

Once a two-process conception is allowed, we can conceive of the controlled attribution process operating as described earlier until a successful resolution of the inconsistency is made, at which point a new trait prototype is used to characterize the stimulus person. Feldman (1972) and Feldman and Hilterman's (1975) stereotyping studies' manipulations may have keyed the process by presenting stimulus persons varying in race, occupation, and social background. Results showed that professional blacks were accorded higher ratings on relevant traits (intelligent, striving, etc.) than comparable whites, supporting the point that unexpected behaviors may elicit a reinterpretation and recategorization of the stimulus person. Consistent or extreme discrepancy from expectations is necessary; a single discrepant behavior may be explained away by a perceiver's causal schemata (Kelley, 1971b), evoking a transient situational attribution. Deaux (1976) reports that women successful at a single masculine task are simply seen as lucky, whereas long-term successes are attributed to effort.

The next section focuses on the question of how information available to the evaluator is integrated into an overall judgment and how perceiver, situation, and task factors influence the process.

Integration and Judgment Processes

In this section, two types of integration processes are considered. The first, described by attribution theory, is cognitive. That is, the final outcome of the process is a categorization or belief statement or a prediction of behavior.

The second integration process is attitudinal or evaluative. It is concerned with the manner in which an overall affective response to an employee is generated from disparate bits of information. Of course, it is true that evaluations and beliefs are related. It is useful to separate these issues,

though, since they are conceptually distinct constructs and (as will be discussed later) different types of formal appraisal instruments call for predominantly evaluative or predominantly cognitive responses, and the determinants of each must be understood.

Cognitive Integration

This topic has already been covered in some detail in the preceding discussion of attribution theory as a controlled categorization process. There, the issue was initial categorization or recategorization of the employee, a situation obviously requiring a summary judgment based on several pieces of information. The present concern is with situations in which similar summary judgments are required for purposes of evaluation or decision making. Deciding which of several potential candidates for higher-level management should be sent to an assessment center or to special executive development programs exemplifies decisions that may require cognitive integration.

The use of "may require" is deliberate. As the foregoing discussion of automatic processes indicates, controlled cognitive integration or attribution takes place only where there is some problem to be resolved. In situations where the automatic categorization and prototype-based inference processes are unchallenged, active integration of information would not take place. Thus, for many years women were thought to be unsuitable in ability or temperament for many positions. Categorization as female would preclude further consideration of the woman's attributes. The same could be said for members of many ethnic or minority groups.

When cognitive integration is required, the processes described earlier are thought to operate. For example, consider a performance review conducted as part of an MBO program. The employee is likely to get systematically greater credit (or blame) than is actually deserved, since people are seen as more causal than they in fact are. The observing supervisor is likely to overestimate the contribution of the employee to any positive or negative results, particularly if he or she belongs to a group not common in the

organization, say women, blacks, or some ethnic minority. To the extent the supervisor is pleased or displeased with the person's or unit's performance, hedonic relevance indicates that stronger attributions to the person will occur. If personal feelings (or prejudice) are involved, biases in the causal attribution of behavior will benefit the liked employee and harm the disliked. Finally, the more poorly defined the objectives were in the first place, the more these biases can influence judgments, since no objective criteria exist against which impressions can be checked. There can be no question as to how many products were produced or sold, for example, but a judgment as to whether maximum effort was exerted is subject to all the biasing factors discussed earlier.

Evaluative Integration

Beliefs about a person are thought to be the basis for the overall attitude toward that person, even though these attitudes may influence subsequent beliefs (Fishbein & Ajzen, 1975). The question of how attitudes are formed, and the precise nature of the function relating belief strength and the evaluative aspect of these beliefs to overall attitude, is the subject of some controversy (see Anderson, 1974, for a general review and further references, and Fishbein & Ajzen, 1975, for a different position). At present, Anderson's weighted-average model, which assumes that new information is averaged with an initial impression of the stimulus person, seems to have the most support. The process is, however, completely dependent on the information available to the evaluator at the time of integration. Salancik and Conway (1975) have shown that relatively simple manipulations can alter the salience of certain information in the recall process, thus changing the overall attitude response (see also Teasdale & Fogarty, 1979). It has already been shown that the process of categorization biases the storage and recall of information and the perceived probability of future behaviors. For example, a good mood may lead to the supervisor's having easier access to positive information about subordinates, raising their average evaluation. A

series of negative events may make negative memories more accessible, lowering evaluations or increasing the perceived negativity of an omission or mistake by a subordinate (these affective responses would also make the subordinates seem more personally responsible for positive or negative actions).

Of course, an evaluative integration may not always occur at the point of judgment. People appear to store overall evaluations separately from other information (Anderson, 1974), and certain questions (e.g., global rating scales) may simply evoke this stored impression. It is also possible that a person associated with several categories (e.g., hard worker, poor dresser, short person) may be characterized by an evaluation associated with each categorization. In such a case, the evaluation associated with the most salient category would be used. If several categories were evoked, the final evaluation would likely be a weighted average of the separate category-based evaluations.

Summary

When decisions about evaluations of employees are necessary, information must be retrieved from memory. Recent research has shown that memories about people (except for their very recent behavior) are biased toward the prototypes representing the categories to which they have been assigned. Dispositional factors in the perceiver and situational influences render certain categories more salient than others and certain memories (and prototype-generated false memories) more available than others. Via the operation of representativeness and availability heuristics in judgment, these further bias the prediction of future behavior. Seeking further information to improve evaluation is an inherently biased process, for people's information-gathering strategies selectively attend to supportive information and elicit hypothesis-confirming behavior.

Under some circumstances, integration of information to form beliefs or evaluations is unnecessary—these are simply recalled as the stored output of earlier integrations. Factors influencing category salience also operate here. Where cognitive integration is necessary, the controlled attribution process described earlier is invoked, and both causal and dispositional attributions result. Evaluative (attitudinal) integration is based on beliefs either recalled or generated by controlled processes and is thought to be described best by a weighted-average combination of component or subsidiary evaluations also affected by salience-causing factors.

A Hypothetical Description of the Appraisal Process

How is all this to be related to the evaluations of performance made in a job setting? To do this, one must consider the task that is to be accomplished.

The first section that follows describes how, using the processes discussed earlier, supervisors are thought to make informal appraisals of their employees—those not requiring a formal appraisal instrument. These would include job assignments, timespan of discretion, extent of supervisory consultation and support, and so on. In short, all those aspects of supervisory behavior that distinguish the *in* employee from the *out* (Dansereau et al., 1975) are considered to be the result of such informal decisions.

The second section below concerns formal appraisals—those requiring the supervisor to complete some official form that becomes part of the employee's record and that may be used as the basis for decisions about pay, promotion, transfer, and so on. Of course, it is explicitly recognized that not all the variance on such forms can be attributed to information processing. They can also be used to justify decisions made previously, not all of which concern merit. Further, the actual matching of recalled behaviors or impressions to given numerical or graphic scales is beyond the scope of this section, which deals only with the information processing engendered by different types of formal evaluation instruments.

Informal Evaluations and Decisions

Setting aside specific individual and organizational policies that may determine ev-

eryday decisions using prespecified objective criteria (Green & Mitchell, 1979), we may ask how the cognitive processes noted above influence such decisions.

If a generalized expectancy model (Campbell et al., 1970) describes the supervisor's choice of alternative behaviors toward subordinates, the appropriate variables to examine are the supervisor's expectancies (subjective probabilities) about the behavior of each subordinate. To generate these, a supervisor might attempt to recall information relevant to the subordinate and situation in question. Information recalled will be of two sorts, behavioral and dispositional (i.e., traits). Both types will be determined by the nature of the prototype representing the category or categories to which the employee in question has been assigned and which are most available for recall and the prototype(s) characterizing the situation or task in question. Further, the memory can be distorted by features of the prototype never actually observed but associated with the subordinate when the category assignment was made. Explicit behavioral information may also be available from the workspace but this information will have been selected by either the automatic or controlled detection processes discussed above. Well-known employees will differ from others chiefly in the variety and amount of information available about them but not in the process by which the information is recalled.

Mischel, Jeffrey, and Patterson (1974) have shown that information about past behavior is preferred for predictions about the behavior of people in situations similar to those in which the original observations took place. When making decisions about work assignments and other familiar tasks, then, the supervisor will attempt to recall specific behaviors. Unless relatively recent relevant behavior has occurred, however, behavioral data must be generated by the elicited category prototype. The representativeness heuristic will operate: behaviors most similar to the category prototype will be seen as most probable.

Mischel et al. (1974) have also shown that people prefer to use trait information when making inferences to situations dissimilar to those in which the person's behavior was originally observed. This trait information is supplied by the relevant prototype and matched to the situational requirements as given by the relevant situational prototype (Bem & Funder, 1978; Wyer & Srull, 1980; Cantor Note 5). Thus, a sales job calls for a "go-getter" whereas bookkeeping demands thoughtfulness and caution. Tversky's (1977) feature-matching process is relevant here. The person that is seen as most likely to be successful will be the one whose perceived traits match the perceived demands of the situation (see Bazerman & Atkin, Note 1).

Formal Evaluations

The most prominent aspect of informal decisions is that they are unstructured; the only requirement is that a set of people must be matched to a set of tasks. But more structured appraisals are also required of supervisors. These are made on formal evaluation forms, where one category or scale point is checked rather than another. The structure of such forms interacts with the processes of categorization, storage, recall, and integration to partially determine the response made (see also Stone & Slusher, 1975). The following discussion considers several frequently used types of formal evaluation instruments, and the cognitive processes thought to underly responses to them.

The simplest type of formal evaluation is a global rating of effectiveness. This instrument requires an evaluative response. The processes described in the attitude literature (Anderson, 1974; Fishbein & Ajzen, 1975) describe such responses well: Available information is combined with some stored impression to produce an overall internal response, which is then mapped onto the scale provided. Both the information available in memory and the evaluative impression itself are outcomes of the processes discussed above; given that general impressions are stored separately from behavioral information, integration need not take place. Simple recall, as moderated by mood or other factors, guides the response. The ranking type of evaluation form, whether based

on paired comparisons among employees or done directly, is similar in that the decision as to which of two employees is more valuable is based on the magnitude of the internal evaluative response in each case. Otherwise the processes are here regarded as identical.

Another commonly used type of evaluation form requires trait ratings about each employee—ability, motivation, reliability, job knowledge, attitude, and so forth. It should be apparent that these are supplied by recall of the relevant prototype in each case, especially if the supervisor has a set of job-relevant prototypes as part of his or her implicit personality theory.

Two interesting implications emerge from this simple concept. First, since sets of traits are part of any prototype, the recall process guarantees a halo effect on trait rating scales. The traits are recalled as a bundle and covary with category just as hand tools and overalls covary with membership in the category skilled worker. The more prototypical the stimulus person, the greater will be the halo effect. Greater knowledge of a person does not circumvent this process. Rather, a well-known person may be assigned to a more explicit category whose prototype differs somewhat from the more general prototypes used for others. Second, leniency/stringency biases can be seen as due in part to the supervisor's cognitive structure or personal construct system. If evaluatively positive categories are habitually more accessible than others, ratings would tend to be lenient; if negative categories are more accessible, ratings would tend to be harsher. Depressive people, for example, show faster recall for unpleasant events (Teasdale & Fogarty, 1979); the availability heuristic would imply that these people would see negative behaviors as more probable. Transient factors, such as mood, operate similarly. Apparently, there is a sound basis for "catching the boss on a good day."

The forced-choice inventory presents somewhat different problems. The advantage of a forced-choice inventory is that behaviors descriptive of the individual must be selected from among a set of equally desirable behaviors. Some of these behaviors discriminate the good from the poor employee,

and others do not; the evaluator cannot deliberately bias the evaluation, and responses are less subject to purely evaluative halo effects. Because the evaluator must choose which behaviors are more likely to be true of the employee, once again the representativeness heuristic is relevant. The behavior in each set most similar to the prototype characterizing the individual is the one most likely to be chosen, whether or not the actual behavior has ever been observed.

The behaviorally anchored rating scale (BARS) requires evaluators to rate each employee on several dimensions of job behavior. Each dimension is defined by several behaviors, scaled to range from exceptionally good to exceptionally poor performance. The rater is presumed to match recalled job behaviors of the employee with those defining the scale points. Thus, the scaled behaviors anchor the judgments. As pointed out by Atkin and Conlon (Note 2), however, many processes may underly judgments on this scale. For example, the rater may search the scale for behaviors matching a general evaluation of the employee. This reduces the BARS to a complex version of the global evaluation scale discussed above. Alternatively, the rater may attempt to seek out behaviors typical of the employee. This process is the same as that discussed for the forced-choice scale and depends on the nature of the prototype characterizing the employee and the operation of the representativeness heuristic. Though no direct evidence on these hypothetical processes exits, some studies are suggestive.

Research has shown (Borman & Dunnette, 1975) that the BARS may be as subject to halo and other sorts of biases as more simple rating scales. This is consistent with the notion that the same processes are involved in the use of BARS as in other scale types. Interestingly, training may decrease several types of bias (Bernardin & Walter, 1977; Borman, 1975; Brown, 1968; Latham, Wexley, & Pursell, 1975; but cf. Warmke & Billings, 1979), suggesting that the training develops categories and associated prototypes more in keeping with actual job requirements; these would allow behavior to be noticed, stored, and recalled in a more useful way. Also relevant is the finding that

more cognitively complex people make better use of BARS than the cognitively simple (Schneier, 1977). Since cognitive complexity is based on the number and relationships among the categories used by an individual (Feldman & Hilterman, 1975), it is not surprising that those with a more differentiated category system would be able to make the relevant judgments with less error.

Behavioral observation scales (Latham & Wexley, 1977) may be subject to prototype-based distortions in much the same way as BARS. The supervisor seeking to recall the frequency with which various behaviors were observed is performing a recognition task similar to that of subjects in memory-bias studies. As such, his or her responses are subject to the same kinds of distortions observed in the laboratory. Of course, the other processes (evaluative matching, selection of typical behaviors, etc.) discussed for the BARS may also occur. The behavioral observation scale may have substantial validity when used as a guide to real-time behavioral observation in a sampling design. Whether or not this is possible depends on organizational factors beyond the scope of the present discussion.

Sources of Bias in Performance Evaluation Reconsidered

It should now be obvious where previous studies of racial, sexual, and cultural biases in performance appraisal have gone astray: They have not taken into account the time-dependent nature of categorization processes nor have they dealt with the problems of selective attention and biased recall. Feldman and Hilterman's (1977) study, for example, probably made employee behavior so salient to the subjects that other variables (race, dress, and employment status) had little influence. Although previous studies such as that one are internally valid, they are not readily generalizable to the everyday employment situation because they do not reproduce the conditions necessary for categorization and integration processes to proceed naturalistically.

Returning to the explicit theoretical consideration of biases in performance appraisal, we may regard stereotypes as sub-classes of prototypes, functionally identical to the trait prototypes of "implicit personality theory" (Ashmore & Del Boca, 1979). Thus, when a stimulus person matches a given racial or ethnic prototype sufficiently well, we may postulate that (at least initially) the automatic process envisioned earlier governs the collection and recall of data about the person. Depending on the nature of the prototype in question, positive or negative consequences may result in terms of job assignments, internal or external attributions for behavior, and the like. Further, the prototype in question will determine the hypotheses that guide any active search for information relevant to the person.

If the person's behavior is observed to be sufficiently discrepant from the expectations generated by the category prototype, a controlled categorization (attribution) process is postulated to occur. As the stereotype research of Feldman (1972) and Feldman and Hilterman (1975) has shown, when long-term behavior contrasts with categorical expectations, a stronger attribution in the direction of contrast is made (corresponding to Jones & Davis's, 1965, notion of "out-of-role" behavior). This attribution presumably provides a new prototype governing selection and recall of information. Thus, the present position predicts both systematic underevaluation or overevaluation of stimulus persons who are members of stereotyped groups provided their behavior either confirms or strongly contradicts their initial categorization and sufficient time passes for selective attention and recall processes to operate.

The complement of this hypothetical sequence is expected to occur for people placed in a positive category, such as the physically attractive individual (Berscheid & Walster, 1974). It may well take more unexpected negative behavior to overturn an initially positive impression in a real-world setting, simply because the initial decisions regarding the individual will be such as to provide greater opportunity for success.

Suggestions for Future Research and Immediate Application

Past research on performance appraisal has been both cognitively and psychometri-

cally oriented. Both traditions should continue but with modifications in both form and content suggested by the foregoing analysis.

On the psychometric side, it should be realized that the search for a best evaluation form is foredoomed unless the processes generating responses to particular scale types are better understood. There seems to be no magic way of phrasing questions that will eliminate response biases, halo effects, leniency/stringency, and so on in the absence of changes in the raters themselves (such as training seems sometimes to produce) or changes in the task (such as requiring behavior sampling). It should also be realized that there is a practical upper limit to the validity (considered as both freedom from bias and accuracy) of any rating form. This limit is set by the reliability of human behavior itself, the reliability of even the best-trained observer/evaluators, transient factors of the situation (such as mood) that cannot be controlled, and the simple fact that most people charged with evaluating others have many other activities that also must be performed and thus a limited amount of time that can be spent on observation and evaluation.

There is, however, a vital contribution that can be made by the psychometrically inclined researcher, one as necessary to the evaluation process as any cognitive theory. The taxonomy of job behavior is in its infancy. At this point, for many jobs (especially at managerial levels) there simply are no good descriptions available. Although substantial progress has been made (e.g., Dowell & Wexley, 1978; McCormick, DeNisi, & Shaw, 1979), much more remains to be done. Especially important is the question of how the dimensionality of job behavior changes as a function of level, technology, and structure.

Description alone is insufficient. To evaluate performance, behaviors must be scaled on some dimension of worth. It is likely that this is a multidimensional problem best approached with a variety of techniques. The development of hard criteria to supplement ratings is another important area of research, one in which the collaboration of accountants, engineers, and other profes-

sionals is necessary. Finally, the appropriate weights and functions for combining multiple criteria into some estimate of the employee's utility to the firm remains to be determined (Cronbach & Gleser, 1965). Of course, these recommendations have been made before. The BARS and behavioral observation scales can be seen as attempts to provide both description and scaling in a single instrument, and these are substantial contributions. The present suggestion is that the work of description and scaling be carried out directly, without immediate concern for the development of an instrument. This should be done in conjunction with other applied research, to be discussed below.

One last point should be made. It seems useful to allow the possibility of organizationally ideographic results of job description and behavior-scaling studies. That is, one should not expect to find every dimension in every level of every organization—each is somewhat unique, and only methods of study may prove to be general.

On the cognitive side, which is really the focus of this article, many recommendations can be made. The first of these is the simplest: stop searching for bias. Bias is an important social problem, but looking for it in a simplistic way will not solve the problem (if it indeed exists in this form). Instead, attention should be focused on two issues: how to represent the evaluation situation realistically in laboratory settings, so that theoretical implications can be tested in an evaluation context, and how to observe and manipulate relevant variables in field settings, so that the findings of the laboratory can be both tested in other settings and put into practice.

For laboratory settings, the following recommendations can be made:

1. The laboratory observer/evaluator cannot be made to focus exclusively on the simulus person's behavior, simply because this does not happen in everyday life. Observations must be made as part of other tasks that the person is doing under some constraint, such as time pressure or a reward/punishment contingency. The stimulus person must be observed emitting an ongoing stream of behavior over a substantial period of time if the results are to be directly

(rather than theoretically) relevant to employment settings.

2. Stimulus persons must vary in the prototypicality of their personal characteristics and behavior. This requires prior studies to define category prototypes.

3. The salience of particular stimulus persons and certain of their prototype-relevant characteristics must be manipulated; for maximum external validity of results, this must be done in ways similar to situations found in the field (e.g., the problems of being a token).

4. Category accessibility should be manipulated in a variety of ways, such as priming or mood induction.

5. Studies must be long term, to allow the effects of categorization on attention and recall to be observed.

6. Measures must not be limited to evaluations but must include means of estimating prototype-based selective attention, free recall, recognition, causal and trait attribution, and the like. Measures such as response latencies must also be considered, especially when studies are designed to investigate the existence of automatic and controlled processes.

7. Studies must be designed to complement psychometric research. For example, what is the effect of rating format on judgment processes? Do some scales make certain categories more accessible than others? With what result? What is the effect of certain kinds of matching or mismatching between task behavior and rating dimensions? The effect of greater or lesser prototype relevance of scales with behavioral content?

8. If bias is studied, it should be studied in the theoretical context given above, and questions of evaluation and attribution must be separated from questions of reward. For example, if a woman is offered a lower salary than a man in a laboratory study, this can mean that the decision maker thinks the woman is worth less or it can mean he or she is trying to maximize economic gains. If a low-performing black is rated higher than a low-performing white, is this rating due to notions of justice, social desirability, or some attributional process?

9. Individual differences must be considered. Ideographic variables such as cognitive structure (personal constructs, implicit personality theory) and nomothetic concepts like cognitive complexity should be included.

10. Interventions must be tested in laboratory settings. It would be helpful to know if a proposed training program influenced the category structure of an individual in a desirable way, and whether this in turn improved the validity of evaluations, prior to that program's being tested in a field setting. Aside from considerations of efficiency, a proper test demands a placebo-trained control group whose evaluations may well be less valid than otherwise, and there are serious ethical problems involved in subjecting actual employees to a deliberately invalid procedure.

For field research, the following recommendations seem useful:

1. Develop measures of cognitive structure, including category systems, prototypes, and accessibility of categories. These would allow researchers to predict what kinds of selective attention and biased recall might occur. In conjunction with an observation and behavior-sampling system applied to work groups, the actual extent of biased recall and selective attention, and their effect on evaluations, could be determined.

2. Using such measures, it would also be possible to determine associations among organizational level, experience, task characteristics, technology, and so on and a supervisor's cognitive representation of employees.

3. The effects of long- and short-term changes in the job and workforce on categorization processes, both automatic and controlled, and on resulting cognitive structures could be studied longitudinally. This would be a useful area for the application of causal modeling techniques. What is the effect, for example, of increasing numbers of women in nontraditional occupations? The effect of new production technology or other job changes?

4. The development of training programs (based on job analyses and scaling studies designed to capture the prototypes and category systems of experienced employees) can best be done in field settings. Likewise, final testing of such programs must be carried out in the field, with attention paid to the issues

covered in paragraphs 1–3 above. Also of interest is the motivational impact of improved evaluations, as reflected in on the job behavior, turnover, absenteeism, and satisfaction measures.

Suggestions for Immediate Application and a Ray of Hope

The reader may have concluded by this time that present performance appraisals are hopelessly biased, and no useful improvement can be made without a long-term (and expensive) research program. This conclusion is premature. Good decisions *are* made in organizations, and before accepting the null hypothesis (that the promotion and reward of the truly competent is a random event), some reasons to expect good decisions should be articulated.

Rosch et al. (1976) assert that categories and their attendant prototypes develop out of observations of the naturally occurring covariation among cues. Given the reasonable assumption that in most cases it is possible to distinguish between employees who are acceptable and those who are utter failures, it would be surprising if experienced evaluators had not at least developed prototypes of good and poor employees. Further, to the extent that the quality of job performance is specifiable in terms of observable behaviors (e.g., parts made to specification, absence, sobriety, courtesy to customers, etc.), the prototype will be valid. This is not to say that no systematic errors will appear; the processes discussed earlier will naturally operate to introduce some systematic error, as in the "illusory correlation" effect (see Hamilton, 1979).

Even where observable behaviors adequate to specify the quality of job performance are lacking, the situation is not completely hopeless. Often, consensual standards exist against which performance can be measured and that serve to guide prototype formation. Since these standards are less open to reliable observation, and are subject to differences in interpretation, less validity should be expected. Still, some (perhaps substantial) validity might be expected in the prototypes of, say, art critics, executive supervisors, journal reviewers, and the like.

Cantor, Smith, French, and Mezzich (Note 6) have, for example, usefully applied prototype concepts to psychiatric diagnosis, an area not known for high reliability.

In short, many evaluation procedures currently in use probably possess some degree of validity, even though they may allow no more than a good–bad judgment. The research strategies suggested above, if successful, would provide more differentiated evaluation procedures, allowing finer discriminations of quality along multiple dimensions of behavior.

Some immediate improvements in evaluation systems can be suggested based on the foregoing analysis. The development of BARS and behavioral observation scales may be seen as an attempt to define a more valid prototype of the successful and unsuccessful employee. Likewise, training in the use of such scales can be seen as an attempt to teach common prototypes to a set of raters. If these efforts were modified to take the fuzzy nature of categories into account and explicitly define more- and less-valued employee prototypes, validity might be improved.

Where possible, multiple evaluations should be obtained. This would tend to overcome the idiosyncratic biases of any one person's category system though retaining the valid aspects of the performance-category prototypes common to both. Of course, this would also retain common biasing factors, but some improvement should result.

The use of hard criteria should be encouraged wherever reasonable. Obviously, for example, punctuality or absenteeism are not appropriate for every job, but where they are, they should be used. The use of job samples as decision tools for promotions and transfers should likewise be encouraged and personal interviews, recommendations, and the like downplayed wherever possible.

Evaluators should be trained to make behavior sampling a routine part of the job to circumvent memory-related biases. Particularly noteworthy behaviors should be recorded for consideration when decisions about employees are to be made. Attentional bias is a potential problem here; perhaps some of the technology employed in behavior modification could be usefully transferred.

Of course, this assumes an explicit definition of what behaviors are relevant or some consensus on the standards of good and poor performance.

The use of trait ratings should be discouraged. In my opinion, whatever validity these may have is probably carried in their evaluative component. They could well be replaced by simpler rating or ranking measures supplemented by employee records (such as lateness or absenteeism).

Finally, the organization must establish an atmosphere in which valid evaluations have some positive consequences for both the evaluator and the evaluatee. The best evaluation system imaginable will be useless unless its consequences are meaningful. If, for example, a poor employee must be given a positive evaluation so he can be transferred, the evaluation system is meaningless. Likewise, if a reward system is so structured that the consequences of a very positive and a merely adequate evaluation are identical, the time spent developing an accurate, unbiased, highly differentiating system has been largely wasted.

It must be remembered that evaluation systems are means, not ends. Larger organizational and societal goals will ultimately determine whether a large-scale research program is justified. It is also true, though, that the usefulness of an innovation is not always realized before the fact. Certainly the attainment of goals such as equal opportunity, more effective selection systems, and better-administered reward systems will be aided by improvements in evaluation. Progress will be made when employers and employees realize that what now seems good enough—isn't.

Reference Notes

1. Bazerman, M. H., & Atkin, R. S. *Performance appraisal: An information-processing and attributional perspective* (Working Paper No. 25-78-79). Pittsburgh, Pa.: Carnegie-Mellon University, Graduate School of Industrial Administration, 1978.
2. Atkin, R. S., & Conlon, E. J. *Behaviorally anchored rating scales: Some theoretical issues* (Working Paper No. 77-75-76). Pittsburgh, Pa.: Carnegie-Mellon University, Graduate School of Industrial Administration, 1977.
3. Smith, E. E., & Medin, D. *Representation and processing of lexical concepts.* Paper presented at the Sloan Conference, University of California, San Diego, Calif., March 1979.
4. Srull, T. K. Personal communication, Spring 1979.
5. Cantor, N. *Perceptions of situations: Situation prototypes and person-situation prototypes.* Paper presented at the conference on the Situation in Psychological Theory and Research, Stockholm, Sweden, June 1979.
6. Cantor, N., Smith, E. E., French, R., & Mezzich, J. Psychiatric diagnosis as prototype categorization (Working Paper). Princeton, N.J.: Princeton University, 1979.

References

Abelson, R. P. Script processing in attitude formation and decision making. In J. S. Carrol & J. W. Payne (Eds.), *Cognition and social behavior.* Hillsdale, N.J.: Erlbaum, 1976.

Anderson, N. H. Cognitive algebra: Integration theory applied to social attribution. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7). New York: Academic Press, 1974.

Ashmore, R. D., & Del Boca, F. K. Sex stereotypes and implicit personality theory: Toward a cognitive-social psychological conception. *Sex Roles,* 1979, *5,* 219–248.

Bell, L. G., Wicklund, T. A., Manko, G., & Larkin, C. When unexpected behavior is attributed to the environment. *Journal of Research in Personality,* 1976, *10,* 316–327.

Bem, D. J., & Funder, D. C. Predicting more of the people more of the time: Assessing the personality of situations. *Psychological Review,* 1978, *85,* 485–501.

Bernardin, H. J., & Walter, C. S. Effects of rater training and diary-keeping on psychometric error in ratings. *Journal of Applied Psychology,* 1977, *62,* 64–69.

Berscheid, E., & Walster, E. Physical attractiveness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7). New York: Academic Press, 1974.

Bigoness, W. J. Effect of applicants' sex, race, and performance on employers' performance ratings: Some additional findings. *Journal of Applied Psychology,* 1976, *61,* 80–84.

Borman, W. C. Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. *Journal of Applied Psychology,* 1975, *60,* 556–560.

Borman, W. C. Exploring the upper limits of reliability and validity in job performance ratings. *Journal of Applied Psychology,* 1978, *63,* 135–144.

Borman, W. C., & Dunnette, M. D. Behavior-based versus trait-oriented performance ratings: An empirical study. *Journal of Applied Psychology,* 1975, *60,* 561–565.

Brown, E. M. Influence of training, method, and relationship on halo effect. *Journal of Applied Psychology,* 1968, *52,* 195–199.

Bruner, J. S. Social psychology and perception. In E. E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* New York: Holt, Rinehart & Winston, 1958.

Campbell, J. P., Dunnette, M. D., Lawler, E. E., III, & Weick, K. *Managerial behavior, performance, and effectiveness.* New York: McGraw-Hill, 1970.

Cantor, N., & Mischel, W. Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology,* 1977, *35,* 38–48.

Cantor, N., & Mischel, W. Prototypes in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 12). New York: Academic Press, 1979.

Cash, T. F., Gillen, B., & Burns, D. S. Sexism and "beautyism" in personnel consultant decision making. *Journal of Applied Psychology,* 1977, *62,* 301–310.

Cronbach, L. J., & Gleser, G. C. *Psychological tests and personnel decisions.* Urbana: University of Illinois Press, 1965.

Dansereau, F., Graen, G., & Haga, W. J. A vertical dyad approach to leadership in formal organizations. *Organizational Behavior and Human Performance,* 1975, *13,* 46–78.

Deaux, K. Sex: A perspective on the attribution process. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 1). Hillsdale, N.J.: Erlbaum, 1976.

Dowell, B. E., & Wexley, K. N. Development of a work-behavior taxonomy for first-line supervisors. *Journal of Applied Psychology,* 1978, *63,* 563–572.

Einhorn, H. J., & Hogarth, R. M. Confidence in judgment: Persistence of the illusion of validity. *Psychological Review,* 1978, *85,* 395–416.

Enzle, M. E., & Schopflocher, D. Instigation of attribution processes by attributional questions. *Personality and Social Psychology Bulletin,* 1978, *4,* 595–599.

Feldman, J. M. Stimulus characteristics and subject prejudice as determinants of stereotype attribution. *Journal of Personality and Social Psychology,* 1972, *21,* 333–340.

Feldman, J. M., & Hilterman, R. J. Stereotype attribution revisited: The role of stimulus characteristics, racial attitude, and cognitive differentiation. *Journal of Personality and Social Psychology,* 1975, *31,* 1177–1188.

Feldman, J. M., & Hilterman, R. J. Sources of bias in performance evaluation: Two experiments. *International Journal of Intercultural Relations,* 1977, *1,* 35–57.

Fishbein, M., & Ajzen, I. *Belief, attitude, intention, and behavior.* Reading, Mass.: Addison-Wesley, 1975.

Garland, H., & Price, K. H. Attitudes toward women in management and attributions for their success and failure in a managerial position. *Journal of Applied Psychology,* 1977, *62,* 29–33.

Gibbs, J. C. The meaning of ecologically oriented inquiry in contemporary psychology. *American Psychologist,* 1979, *34,* 127–140.

Graen, G. Role making processes within complex organizations. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology.* Chicago: Rand McNally, 1976.

Green, S. G., & Mitchell, T. R. Attributional processes of leaders in leader-member interactions. *Organiza-tional Behavior and Human Performance,* 1979, *23,* 429–458.

Hamilton, D. L. A cognitive-attributional analysis of stereotyping. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 12). New York: Academic Press, 1979.

Hamner, W. C., Kim, J. S., Baird, L., & Bigoness, W. J. Race and sex as determinants of ratings by potential employers in a stimulated work-sampling test. *Journal of Applied Psychology,* 1974, *59,* 705–711.

Hastie, R. Memory for information that is congruent or incongruent with a conceptual schema. In E. T. Higgins (Ed.), *Social cognition.* Hillsdale, N.J.: Erlbaum, 1978.

Hastorf, A. H., Schneider, D. J., & Polefka, J. *Person perception.* Reading, Mass.: Addison-Wesley, 1970.

Jones, E. E. The rocky road from acts to dispositions. *American Psychologist,* 1979, *34,* 107–117.

Jones, E. E., & Davis, K. E. From acts to dispositions. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press, 1965.

Jones, E. E., & McGillis, D. Correspondent inferences and the attribution cube: A comparative reappraisal. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 1). Hillsdale, N.J.: Erlbaum, 1976.

Kelley, H. H. Attribution in social interaction. In E. E. Jones et al. (Eds.), *Attribution: Perceiving the causes of behavior.* Morristown, N.J.: General Learning Press, 1971. (a)

Kelley, H. H. Causal schemata and the attribution process. In E. E. Jones et al. (Eds.), *Attribution: Perceiving the causes of behavior.* Morristown, N.J.: General Learning Press, 1971. (b)

Kelly, G. A. *A theory of personality: The psychology of personal constructs.* New York: Norton, 1955.

Kruglanski, A. W. Attributing trustworthiness in supervisor-worker relations. *Journal of Experimental Social Psychology,* 1970, *6,* 214–232.

Langer, E. J. Rethinking the role of thought in social interaction. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 2). Hillsdale, N.J.: Erlbaum, 1978.

Latham, G. P., & Wexley, K. N. Behavioral observation scales for performance appraisal purposes. *Personnel Psychology,* 1977, *30,* 255–268.

Latham, G. P., Wexley, K. N., & Pursell, E. D. Training managers to minimize rating errors in the observation of behavior. *Journal of Applied Psychology,* 1975, *60,* 550–555.

Markus, H. Self-schemata and processing information about the self. *Journal of Personality and Social Psychology,* 1977, *35,* 63–78.

McArthur, L. Z., & Post, D. L. Figural emphasis and person perception. *Journal of Experimental Social Psychology,* 1977, *13,* 520–535.

McCloskey, M., & Glucksberg, S. Decision processes in verifying category membership statements: Implications for models of semantic memory. *Cognitive Psychology,* 1979, *11,* 1–37.

McCormick, E. J., DeNisi, A. S., & Shaw, J. B. Use of the Position Analysis Questionnaire for establishing the job component validity of test. *Journal of Applied Psychology*, 1979, *64*, 51–56.

Medin, E. L., & Schaffer, M. M. Context theory of classification learning. *Psychological Review*, 1978, *85*, 207–238.

Meehl, P. E. Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 1978, *46*, 806–834.

Mintzberg, H. *The nature of managerial work*. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Mischel, W., Jeffery, K. M., & Patterson, C. J. The layman's use of trait and behavioral information to predict behavior. *Journal of Research in Personality*, 1974, *8*, 231–242.

Mitchell, T. R., Green, S. G., & Wood, R. E. An attributional model of leadership and the poor-performing subordinate: Development and validation. In B. Staw & L. Cummings (Eds.), *Research in organizational behavior* (Vol. 3). Greenwich, Conn.: JAI Press, 1981.

Newtson, D. Foundations of attribution: The perception of ongoing behavior. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 1). Hillsdale, N.J.: Erlbaum, 1976.

Nisbett, R. E., & Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 1977, *84*, 231–259.

Regan, D. T., Strauss, E., & Fazio, R. Liking and the attribution process. *Journal of Experimental Social Psychology*, 1974, *10*, 385–397.

Rosch, E. Human categorization. In N. Warren (Ed.), *Studies in cross-cultural psychology* (Vol. 1). New York: Academic Press, 1977.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. Basic objects in natural categories. *Cognitive Psychology*, 1976, *8*, 382–439.

Rosen, B., & Jerdee, T. H. Influence of sex-role stereotypes on personnel decisions. *Journal of Applied Psychology*, 1974, *59*, 9–14.

Rosenberg, S., & Sedlak, A. Structural representations of implicit personality theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.

Ross, L. The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10). New York: Academic Press, 1977.

Ross, L., Greene, D., & House, P. The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 1977, *13*, 279–301.

Salancik, G. R., & Conway, M. Attitude inferences from salient and relevant cognitive content about behavior. *Journal of Personality and Social Psychology*, 1975, *32*, 829–840.

Schneier, E. C. Operational utility and psychometric characteristics of behavioral expectations scales: A cognitive reinterpretation. *Journal of Applied Psychology*, 1977, *62*, 541–548.

Schneider, W., & Shiffrin, R. M. Controlled and au-

tomatic human information processing: I. Detection, search, and attention. *Psychological Review*, 1977, *84*, 1–66.

Shank, R., & Abelson, R. P. *Scripts, plans, goals, and understanding*. Hillsdale, N.J.: Erlbaum, 1977.

Shaver, K. G. *An introduction to attribution processes*. Cambridge, Mass.: Winthrop, 1975.

Shiffrin, R. M., & Schneider, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 1977, *84*, 127–190.

Smith, P. C. Behavior, results, and organizational effectiveness: The problem of criteria. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology*. Chicago: Rand McNally, 1976.

Smith, E. E., Adams, N., & Schorr, D. Fact retrieval and the paradox of interference. *Cognitive Psychology*, 1978, *10*, 438–464.

Smith, E. R., & Miller, F. D. Attributional information processing: A response time model of causal subtraction. *Journal of Personality and Social Psychology*, 1979, *37*, 1723–1731.

Snyder, M., & Cantor, N. Testing hypotheses about people: The use of historical knowledge. *Journal of Experimental Social Psychology*, 1979, *15*, 330–342.

Snyder, M., & Swann, W. B., Jr. Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology*, 1978, *36*, 1202–1212. (a)

Snyder, M., & Swann, W. B., Jr. Behavioral confirmation in social interaction: From social perception to social reality. *Journal of Experimental Social Psychology*, 1978, *14*, 148–162. (b)

Snyder, M., & Uranowitz, S. W. Reconstructing the past: Some cognitive consequences of person perception. *Journal of Personality and Social Psychology*, 1978, *36*, 941–950.

Srull, T. K., & Wyer, R. S., Jr. The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology*, 1979, *37*, 1660–1672.

Stone, T. H., & Slusher, E. A. Attributional insights into performance appraisal. *JSAS Catalog of Selected Documents in Psychology*, 1975, *5*, 253. (Ms. No. 964)

Strickland, L. S. Surveillance and trust. *Journal of Personality*, 1958, *26*, 200–215.

Tajfel, H. Social and cultural factors in perception. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 3). Reading, Mass.: Addison-Wesley, 1969.

Taylor, S. E., & Fiske, S. T. Salience, attention, and attributions: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11). New York: Academic Press, 1978.

Taylor, D. M., & Jaggi, V. Ethnocentrism and causal attribution in a South Indian context. *Journal of Cross-Cultural Psychology*, 1974, *5*, 162–171.

Teasdale, J. D., & Fogarty, S. J. Differential effects of induced mood on retrieval of pleasant and unpleasant events from episodic memory. *Journal of Abnormal Psychology*, 1979, *88*, 248–257.

Terborg, J. R., & Ilgen, D. R. A theoretical approach

to sex discrimination in traditionally masculine occupations. *Organizational Behavior and Human Performance,* 1975, *13,* 352–376.

Triandis, H. C. Cultural influences upon cognitive processes. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 1). New York: Academic Press, 1964.

Tsujimoto, R. N. Memory bias toward normative and novel trait prototypes. *Journal of Personality and Social Psychology,* 1978, *36,* 1391–1401.

Tsujimoto, R. N., Wilde, J., & Robertson, D. R. Distorted memory for exemplars of a social structure: Evidence for schematic memory processes. *Journal of Personality and Social Psychology,* 1978, *36,* 1402–1414.

Tversky, A. Features of similarity. *Psychological Review,* 1977, *84,* 327–352.

Tversky, A., & Kahneman, D. Belief in the law of small numbers. *Psychological Bulletin,* 1971, *76,* 105–110.

Tversky, A., & Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science,* 1974, *185,* 1124–1131.

Warmke, D. L., & Billings, R. S. Comparison of training methods for improving the psychometric quality of experimental and administrative performance ratings. *Journal of Applied Psychology,* 1979, *64,* 124–131.

Weiner, B. *Theories of motivation: From mechanism to cognition.* Chicago: Markham, 1972.

Wyer, R. S., Jr., & Srull, T. K. 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 on personality and social psychology.* Hillsdale, N.J.: Erlbaum, 1980.

Zadny, J., & Gerard, H. B. Attributed intentions and informational selectivity. *Journal of Experimental Social Psychology,* 1974, *10,* 34–52.