INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0500
Are four heads better than one? Comparing groups and individuals on behavioral rating accuracy

Borg, Maria Rita, M.A.
Rice University, 1991
RICE UNIVERSITY

ARE FOUR HEADS BETTER THAN ONE?
COMPARING GROUPS AND INDIVIDUALS
ON
BEHAVIORAL RATING ACCURACY

by

Maria R. Borg

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
MASTER OF ARTS

APPROVED, THESIS COMMITTEE

Dr. R. Martell, Assistant Professor
Department of Psychology

Dr. R. Dipboye, Professor
Department of Psychology

Dr. D. Schneider, Chairman &
Professor Department of
Psychology

Houston, Texas

April, 1991
Are Four Heads Better Than One? Comparing Individual and Group Performance on Behavioral Rating Accuracy

Maria R. Borg

Abstract

The main objective of this research was to determine whether differences between group and individual accuracy on behavioral rating tasks are due to differences in memory sensitivity or to systematic differences in the type of decision criterion adopted. Group vs. individual differences in evaluative judgment and in confidence levels, and the effects of a five-day delay were also investigated. Lastly, the relationship between response bias and prior evaluative judgment was explored. The results revealed a group memory superiority but also demonstrated that groups adopt a too-liberal decision criterion when rating the occurrence of effective behaviors. In addition, in the delayed rating condition, groups were found to be more confident in their correct responses than individual subjects. And finally, for individual subjects, prior evaluative decisions were positively related to response bias in the rating of effective behaviors and negatively related to response bias in the rating of ineffective behaviors.
Acknowledgements

I'd like to thank Dr. R. Martell for providing invaluable guidance and patience throughout this project, and my other committee members Dr. D. Schneider and Dr. R. Dipboye for reviewing this manuscript.

I am also grateful to Marnie Swerdlin for her proof-reading help and to Cindy Willis for her assistance in data analysis.
# Table of Contents

**Introduction**
- Behavioral Rating Accuracy 2
- Practical Implications of Behavioral Rating Accuracy 8
- Evaluative Judgment 9
- Time Delay Effects on Accuracy 10
- Confidence Levels 11
- Goals and Overview 12

**Method**
- Subjects and Design 13
- Stimulus Material 13
- Procedure 14
- Independent Variables 15
- Dependent Measures 16

**Results**
- Evaluative Ratings 17
- Behavioral Ratings 17
- Relationship Between Evaluative Judgment and Decision Criterion 23
- Confidence Level Analysis 23

**Discussion**
- Evaluative Judgment 30
- Memory Sensitivity 30
- Response Bias 32
- Evaluative Judgment and Response Bias 34
- Levels of Confidence 35
- Practical Ramifications 38

**References** 43
## Table of Contents, Continued

<table>
<thead>
<tr>
<th>Footnotes</th>
<th>49</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Appendices</strong></td>
<td></td>
</tr>
<tr>
<td>(A) Vignette</td>
<td>50</td>
</tr>
<tr>
<td>(B) Instructions</td>
<td>58</td>
</tr>
<tr>
<td>(C) Distractor Task</td>
<td>63</td>
</tr>
<tr>
<td>(D) Evaluative Task Materials</td>
<td>65</td>
</tr>
<tr>
<td>(E) Behavioral Rating Task Materials</td>
<td>69</td>
</tr>
</tbody>
</table>
List of Tables

<table>
<thead>
<tr>
<th>No.</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mean Work Performance Rating</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>Mean Hit Rate, False Alarm Rate, Memory Sensitivity and Response Bias</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>Simple Correlations Between Evaluative Rating and Response Bias for Effective and Ineffective Behaviors</td>
<td>24</td>
</tr>
<tr>
<td>4</td>
<td>Mean Number of &quot;Positive&quot; Responses for Different Experimental Conditions</td>
<td>26</td>
</tr>
<tr>
<td>5</td>
<td>Mean Confidence Scores for Correct and Incorrect Responses in Different Experimental Conditions</td>
<td>28</td>
</tr>
</tbody>
</table>
Are Four Heads Better Than One? Comparing Groups And Individuals On Behavioral Rating Accuracy

Maria R. Borg

Business organizations and other social institutions are increasingly depending on judgments and decisions made by groups of interacting individuals. Committees, task forces, selection panels, and assessment centers are only some of the more common types of decision-making groups often called upon to evaluate and "process" various types of information and then arrive at collective judgments and decisions. This increasing dependence on collective judgments in organizations reflects a general belief that "two heads are better than one" and that, the more numerous the intellectual resources the better the quality, or accuracy of the judgments and decisions that result from group interaction. This assumption, however, is only partially supported by empirical research. When groups and individuals are compared on problem solving, decision making or memory tasks, group performance is not consistently superior or more accurate than its individual counterpart (Hill, 1982). One kind of decision process that has often been overlooked in this type of research is the performance appraisal process. Although there has been a surge of interest, during the last decade in the cognitive processes that underlie the assessment of work behavior, this interest has focused almost exclusively on the individual appraiser. Little is yet known about how performance ratings rendered by panels or committees differ from those rendered by individual raters. The main purpose of this research, then, was to address the issue of group vs. individual accuracy on behavioral rating tasks.
A behavioral rating task requires the rater to decide whether each behavior on a list of behaviors (sometimes called a "behavioral checklist") was previously demonstrated by the worker whose performance is being appraised. A basic assumption underlying this study's design is that accuracy on this type of rating task can be decomposed into two separate components, memory sensitivity and response bias. This assumption is a reflection of Lord's (1985) perspective on behavioral rating accuracy. Lord's views were adopted as the conceptual guide for this study and they are briefly discussed next.

**Behavioral Rating Accuracy**

Calling accurate behavioral measurement an "elusive objective," Lord argues that typical measures of accuracy do not adequately detect systematic errors in raters' judgments. Behavioral ratings may reflect not only the sensitivity of the raters' behavioral memory per se (what we usually imply by accurate behavioral measurement, which is simply the ability to distinguish between behaviors that previously occurred and those that did not), but also changes in the decision criteria that raters set for themselves when faced with a recognition-type task, such as a behavioral checklist. Lord uses premises from signal detection theory to propose an approach to behavioral rating accuracy that differentiates conceptually and operationally between memory sensitivity and systematic changes in decision criteria. Hit rates (HR, the proportion of "old" behaviors correctly recognized) and false-alarm rates (FAR, the proportion of behaviors incorrectly recognized as "old"). HR and FAR which are operationally related to strengths of
signal plus noise and noise alone (respectively), can be translated into a measure of memory sensitivity. Besides pure discriminability, HR and FAR are also affected by raters' decision criteria; an estimation of the leniency or stringency of raters' mode of responding yields a measure of response bias that is conceptually and operationally independent of the memory sensitivity measure. The main interest of this study was to investigate group vs. individual differences on behavioral rating tasks in terms of some combination of memory sensitivity and response bias. Lord's conceptualization was ideal because it allowed for the examination of behavioral ratings in terms of both processes. By using measures of memory sensitivity (Pr) and response bias (Br) derived from a "two-high threshold model" of recognition memory (see Snodgrass & Corwin, 1988), the present study was able to take a rigorous approach to measuring differences in accuracy between group and individual ratings. Because the Pr measure corrects for guessing, it was possible to separate true accuracy in recognizing previously presented behaviors from indiscriminate guessing. Additionally, the Br measure, made it possible to determine whether when unsure, groups adopted a too-liberal (bias to say a behavior did occur), or a too-conservative (bias to say a behavior did not occur) decision criterion in relation to individuals.

The few studies that have utilized similar methods in the context of performance appraisal have involved only individual raters (see Murphy, Philbin & Adams, 1989; Martell & Guzzo, in press; Martell & Willis, 1991). Comparisons of group and individual
performance on tests of recognition memory are also infrequent and when they are made, the general conclusion seems to be that groups tend to outperform the individual subject. A recent study by Vollrath, Sheppard, and Hinsz (1989) attempted to generalize beyond the simple stimuli (nonsense syllables, words, numbers and stories) used in past group vs. individual performance literature and to provide a comprehensive comparison of group and individual processing of complex behavioral information. Subjects listened to a pre-recorded mock trial featuring 25 chronologically arranged behavioral facts and later performed a behavioral recognition task individually or collectively in groups of four. Groups were required to come to a consensus decision in deciding whether or not each behavior had in fact occurred in the "trial". As hypothesized, it was found that groups correctly identified a higher percentage of items (higher Hit Rate) than did individuals working alone and were thus assumed to be more accurate. In a study by Hartwick, Sheppard and Davis (mentioned in Guzzo, 1982) also involving mock jurors, subjects viewed a videotaped trial and later performed a recognition task in groups of four or individually. Again, as predicted, groups outperformed individuals in terms of the greater number of correctly identified items.

These findings not only have significant theoretical implications for the nature of group process, but also important practical implications for the use of behavioral checklists in performance appraisal. They suggest that a group of individuals collectively using a behavioral checklist, which is essentially a recognition task, can produce a more accurate performance appraisal of a subordinate
because they can remember more of his or her on-the-job-behaviors. However, these studies comparing group and individual performance may not have examined behavioral rating accuracy rigorously enough to justify a confident statement about the superiority of group performance. In both of the studies mentioned above, there was no correction for guessing, and differences in decision criteria were not measured. The higher hit rate could also have been a function of simple guessing and/or a more "liberal" decision criterion adopted by the group members when they were unsure about how to respond. When deciding whether a presented item on a recognition task is an old or a new one, individuals in a group situation may tend to respond in a specific manner independently of what they actually remember. Collectively they may, for instance, more readily report that items are old rather than new; that is, they may adopt a "too-liberal" decision criterion when they unsure. This type of systematic response bias would then result in a higher hit rate, whereby a higher number of previously presented items are correctly identified. If such is the case, superiority in performance should not be interpreted as a greater capacity for accuracy in such tasks. In fact, Hartwick et al. found that groups were also more likely to claim that an item had been previously presented than to claim that it was a new item. This, of course, contributed to the higher number of correct responses by groups but does not in itself represent greater accuracy in behavioral rating.

Hasty conclusions about the superiority of group memory are further discouraged by the findings of Stephenson, Brandstatter &
Wagner (1983). In this study, individual and dyadic recall of a story - "The War of the Ghosts" - was compared. It was found that although dyads were more frequently correct than individuals, they were also more likely to accept erroneous suggestions and were overconfident in both their correct and incorrect answers. In a more recent study by Stephenson, Clark & Wade (1986), subjects listened to an audiotape of a simulated police interrogation and later were required to reconstruct the interrogation in writing and to answer specific questions related to it. It was found that groups of four individuals who reached consensus decisions about how to respond on both tasks were even more confident in their answers than dyads. Especially noteworthy was that, when groups of four answered incorrectly, their level of confidence was what the researchers called "outrageously high" (p. 1118). The focus in the Stephenson et al. studies was recall and not recognition memory, but nevertheless, these findings do seem to suggest that there is more to accuracy than the sheer volume of material recognized. Furthermore, the overconfidence shown by groups in the Stephenson et al. studies closely parallels the concept of a "freer", or a too-liberal manner of responding to items on a behavioral recognition task.

Why, then, might groups be more apt to adopt a too-liberal response mode when they are uncertain? Consensus decision-making situations, such as the one used in this study, are likely to generate a fair amount of social pressure on individual group members, and this could be one of the factors that affects the type of decision criterion adopted by groups. Despite the fact that groups are told that all
members must agree on a group rating, it is more likely that there is an acutely felt pressure to conform to what seems to be the predominant group judgment. From a group polarization perspective (Myers & Lamm, 1976), group interaction or the social pressure generated therein is likely to magnify the tendency for a slightly biased decision criterion that may already be present in individuals.

The possibility of higher propensity for bias in groups does not eliminate the possibility that groups may indeed be more accurate on behavioral rating tasks. When guessing is controlled for, and the effects of response bias are taken into consideration, group performance on these tasks (in terms of memory accuracy) may still exceed that of individuals. On the other hand, it may only appear that "two or more heads are better than one" in this case; group membership may cause individuals to answer items in a specific manner independently of what they actually remember. The apparent memory superiority of groups may be concealing an underlying response bias that renders collective behavioral rating no more, or maybe even less, accurate than that of individual raters and it was this very issue that the present study sought to investigate. Is the alleged superiority of group over individual performance on these types of tasks due to true differences in memory accuracy, to a systematic response bias whereby unsure groups adopt a too-liberal decision criterion, or maybe to some combination of both processes?

In an effort to provide answers to the above questions, the main objective of this research effort was to re-examine group vs. individual differences on behavioral rating tasks using a more precise measure of
accuracy than that used previously in this type of research.

**Practical Implications of Behavioral Rating Accuracy**

Further knowledge about the exact nature of the "accuracy advantage" that groups are reported to have on these types of judgment tasks is important not only in terms of the development of theories of social information processing but also in terms of the many organizational practices where such processing is involved. The most salient of these, of course, is performance appraisal in the form of behavioral checklists that require raters to decide whether or not different behaviors were displayed by a particular employee. In order to be accurate, such a measure requires not only that the raters "recognize" behaviors that actually occurred, but also that they confidently identify the ones that did not. In this context, it is clear to see why the issue of a systematic response bias and its effects becomes increasingly important.

Another issue that comes up in terms of behavioral checklists is the fact that these typically consist of some combination of positive and negative behavioral examples. It is conceivable that there could be differential effects of bias on the recognition of these two types of behavior. If, in fact, group judgement is "liberal", does it tend to be more so when the group is considering effective behaviors or when it is considering ineffective behaviors? Put another way, do raters use different decision criteria when rating (or recognizing) effective and ineffective behaviors, and does this tend to happen more or less when the task is undertaken collectively by a group of people? It was deemed important, therefore, to analyze memory sensitivity and
response bias separately for effective and ineffective behavioral items, thus paving the way for more direct investigation in future research.

**Evaluative Judgment**

In addition to a behavioral rating task, this study also included an evaluative task that required subjects to make evaluative judgments about an individual's performance. The most fundamental reason for including this evaluative task was to see whether groups tend to evaluate performance more or less favorably than individuals. Early studies on the distorting influence of audiences (Allport, 1924) showed that subjects in a group situation tend to give more moderate judgments, presumably to minimize the possibility of being extremely different from others. On the basis of these findings, it is conceivable that consensus group judgment may result in performance evaluations that are closer to "average" than those made by individual subjects. On the other hand, past findings of group polarization and "risky shift" suggest that groups tend to make decisions that are more extreme than the average of the decision preferences of the individuals making up the group (Pruitt, 1971; Wallach, Kogan, & Bem, 1962). From this point of view, groups may exaggerate individual rating tendencies of leniency and severity, thus being more extreme in their ratings than individuals.

The inclusion of an evaluative measure also allowed for the detection of possible underlying relationships between prior evaluative decisions and response bias. Are favorable performance ratings related to a more liberal decision criterion in the attribution of effective behaviors and to a more conservative decision criterion in the
attribution of ineffective behaviors? In previous studies (Martell & Guzzo, in press; Martell & Willis, 1991), subjects given "positive performance cues" later adopted a more liberal criterion when judging the occurrence of effective behaviors. In the present study, all subjects first evaluated the performance of a police officer (based on a written account of his work behavior). Subjects could have become committed to a general impression of the police officer's performance that they formed before the behavioral rating task. Unlike in the previous studies, this impression would act as an "internal" performance cue. There is no reason to believe that this would be any less powerful in its relationship to raters' decision criteria than an external performance cue. Unfortunately, the direction of the evaluative ratings could not be experimentally controlled in this study and only correlational data were available.

**Time Delay Effects on Accuracy**

Lastly, the present study addressed an issue that has not yet been tackled by research in this area. A delayed rating condition was included in order to examine whether a delay of a few days between the presentation of behavioral stimuli and behavioral rating would moderate the pattern of differences between group and individual performance. The manner in which recollections of behavior and decisions based on those recollections are modified by the passage of time is a matter of considerable practical relevance. With the exception of specialized appraisal procedures such as assessment centers, the assessment of employee performance does not immediately follow behavioral observation. What is more typical is that
a manager or a supervisor is faced with the momentous task of completing performance appraisals on all subordinates at the end of a year. Despite recommendations (Bernardin & Walter, 1977; Hakel, Appelbaum, Lyness, & Moses, 1983) that a behavioral diary method be used as a possible strategy for facilitating the appraisal process, it is doubtful that this method is widely used; thus, raters have to rely on their memory in filling out behavioral checklists on each employee. Because there are likely to be serious gaps in stored recollections of an employee's work behavior, and the remaining information cannot be readily accessed, a "heuristic" may be adopted to facilitate the decision-making process. This heuristic may take the form of a systematic response where the rater, when not certain of the right response, is either too "liberal", saying that most behaviors on a checklist were demonstrated by the ratee, or too "conservative", saying that most behaviors on a checklist were not demonstrated by the ratee.

Confidence Levels

The Stephenson et al. studies (1983, 1986) mentioned above, suggest that confidence is somehow implicated in group performance. Generally, as in the Stephenson et al. studies, confidence ratings are obtained separately from the measures of group performance. Fortunately, recognition measures such as the one used in this study are well suited to the analysis of confidence. A yes-no recognition procedure can be easily extended to obtain a measure of how certain the subject is that a given item was or was not previously presented (Srull, 1984). Because subjects responded to the behavioral rating
task by using a 6-point scale with endpoints labeled (1) "very certain the behavior did not occur" to (6) "very certain the behavior did occur", measuring differences in confidence levels among experimental groups proved to be relatively straightforward.

**Goals and Overview**

To summarize, the main objective of this research was to explore possible group vs. individual differences in behavioral rating accuracy. A rigorous measure of accuracy was utilized in order to determine whether any differences between group and individual performance are due to a true memory difference, a systematic change in decision criteria or, perhaps, both. Another goal was to discover whether in terms of memory sensitivity and response bias, the pattern of results for the rating of effective behaviors differs from that of ineffective behaviors and whether any of the effects are moderated by a time delay. In addition, group vs. individual evaluative judgment was compared by means of an evaluative task and lastly, measures of confidence levels were obtained. In terms of the relationship between the evaluative rating and the decision criterion used in the behavioral rating, it was predicted that subjects who evaluated the police officer favorably would be more likely to adopt a liberal decision criterion when attributing effective behaviors than when attributing ineffective behaviors to that officer. No definite predictions were made in relation to differences between groups and individuals in this respect. As for confidence ratings, it was hypothesized, based on the findings of Stephenson et al., that groups would be more confident in their ratings than individual subjects. In addition to this, it was expected
the subjects' certainty about their own responses making overall at a delay between stimulus presentation and rating would decrease confidence for delayed ratings lower than that for immediate ratings.

Subjects in this study were presented with a written vignette depicting the work behavior of a police officer over a three-day period. Subjects then performed an evaluative rating task and a behavioral rating task based on the presented material either immediately or five days later. Subjects performed the tasks either individually or collectively in groups of four.

Method

Subjects and Design

One hundred and ninety-one Rice University students each receiving course credit, voluntarily participated in the experiment either as individuals or in groups of four. They were randomly assigned to the experimental conditions; their informed consent was obtained prior to the experiment and they were fully debriefed at its conclusion. The design was a $2 \times 2$ factorial with the independent variables being Subject Condition (group or individual), and Time of Rating (immediate and delayed).

Stimulus Material

Subjects read a four-page typewritten vignette about a police officer (see Appendix A). This vignette depicted a police officer's behavior over a three day period and contained both effective and ineffective work behaviors. Before subjects read the vignette they also read a short job-description about a police officer's job and some biographical information about the police officer.
Behavioral incidents contained in the vignette were selected from a pool of a hundred critical incidents generated by police officers (DeCotis, 1974; Landy, 1973; Landy, Farr, Saal & Freytag, 1976). These had been rated as effective or ineffective by 29 independent judges from a similar population using a nine-point rating scale, with endpoint anchors of "absolutely ineffective" and "absolutely effective". The vignette contained 10 effective and 6 ineffective behaviors.

Procedure

The study was introduced to subjects as an organizational simulation in which they were to "...read the vignette with the goal of evaluating the individual's work performance" (see Appendix B for verbatim instructions to subjects). Only one subject condition was run at each experimental session. Subjects were given 15 minutes to read the vignette (pilot testing established that this was sufficient time). To control for differential effects due to the mere presence of others (Zajonc, 1965), all subjects read the vignette in non-interacting groups. Immediately after this, they started working individually on a five-minute distractor task that was meant to prevent active rehearsal of the information in the vignette. This task required subjects to write short answers in response to questions about college life (see Appendix C). Immediately after the distractor task or five days later, subjects completed the evaluative (see Appendix D) and the behavioral rating tasks (see Appendix E). Subjects completed the tasks individually or in groups of four individuals. In order to prevent systematic differences in encoding strategies (see Srull & Wyer, 1983), subjects were not informed until after completion of the
distractor task whether they would be completing the experiment that day or in a few days and whether they would be working alone or as a group.

**Independent Variables**

**Subject Condition: Individual vs. Group.** Subjects worked on the evaluative and behavioral rating tasks either individually - **Individual Condition** - or in groups of four individuals - **Group Condition**. In both conditions, subjects had 5 minutes to complete the evaluative task. After this, they were given 25 minutes to complete the behavioral rating task (again, it was previously established that this was sufficient time to complete ratings). For the evaluative task, groups were asked to read the items and come to a consensus decision about what rating to give the police officer on each item. It was stressed that all members must agree on the rating before it was recorded by one of the group members. On the behavioral rating task, groups were instructed to reach a consensus decision about whether or not that behavior occurred in the vignette and to record their decision on appropriate level of the scale.

**Time of Rating: Immediate vs. Delayed.** There were two different times of rating: an **immediate** rating and a **delayed** rating. In the immediate condition, subjects worked on the evaluative task and then the behavioral rating task immediately after the five minute distractor task. In the delayed condition, subjects were released immediately after the distractor task and instructed to return in five days' time in order to complete the experiment and receive course credit. On that date, they worked on the tasks as in the immediate
condition. Again, in order to prevent systematic differences in encoding strategies, subjects did not know until they returned whether they would be working individually or in a group.

**Dependent Measures**

**Evaluative Ratings.** Four 9-point graphic rating scales were used to evaluate the police officer's competence ("very competent-not very competent"), job performance ("excellent-poor"), potential for advancement ("very favorable-not very favorable"), and future success ("very successful-not very successful"). Responses on this measure were examined for differences in leniency and severity. That is, it was determined whether the officer was rated more positively or more negatively in certain experimental conditions.

**Behavioral Ratings.** Subjects' recollections of the police officer's behavior were assessed with a 32-item questionnaire. Using a 6-point scale with endpoints labeled (1) "very certain the behavior did not occur" to (6) "very certain the behavior did occur", subjects indicated whether the behaviors did or did not occur in the vignette. Sixteen of the 32 items depicted behaviors which actually occurred in the vignette and another 16 depicted behaviors that did not occur. Of the 16 "true" behaviors, 10 were "effective" in that they described behavior reflecting favorably on the police officer, and 6 were "ineffective" in that they described behavior reflecting unfavorably on the police officer. From responses on this measure it was possible to determine how many behaviors were correctly recognized (a measure of memory sensitivity after correcting for guessing) and whether there were differences in the decision criterion adopted by subjects in different
conditions (a measure of response bias). A method recommended by Snodgrass & Corwin (1988) in their recent review of measures of memory sensitivity and response bias was used to analyze responses on this measure.

Results

The Type of Subject and Time of Rating manipulations had no effects on the evaluative ratings. However, they did have significant effects on the behavioral ratings, the relationship between evaluative judgement and decision criterion, and the confidence levels.

Evaluative Ratings

The evaluative ratings were analyzed in a $2 \times 2$ (Subject Type: Individual/Group x Time of rating: immediate/delayed) between group analysis of variance.

Work Performance. A mean Work Performance score was created by averaging subjects' ratings of the police officer's competence, job performance, potential for advancement and likely future success (coefficient alpha = .88). The means appear in Table 1. Analysis of variance did not produce significant main effects or interactions. Contrary to predictions, individuals and groups did not differ in their evaluative ratings and immediate ratings were not significantly different from those made after a delay of five days.

Behavioral Ratings

The first task was to determine whether groups and individuals differed in the proportion of work behaviors they attributed to the police officer and thus the overall hit rates for each individual and each group were calculated. This was done by treating behaviors that
Table 1
Mean Work Performance Ratings

<table>
<thead>
<tr>
<th>Subject Condition</th>
<th>Immediate</th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td>\text{Individuals}</td>
<td>5.31</td>
<td>5.57</td>
</tr>
<tr>
<td>\text{(sd=1.74)}</td>
<td>32</td>
<td>23</td>
</tr>
<tr>
<td>\text{Groups}</td>
<td>5.84</td>
<td>5.30</td>
</tr>
<tr>
<td>\text{(sd=1.47)}</td>
<td>19</td>
<td>15</td>
</tr>
</tbody>
</table>

A 9-point Likert-type scale was used. A rating of 1 was the least favorable rating, 5 was neutral, and 9, was the most favorable rating.
occurred in the vignette and rated a 4, 5, or 6 (indicating some
degree of confidence) as "hits," while treating behaviors not occurring
in the vignette but mistakenly rated as a 4, 5, or 6 as "false alarms."
Subjects' overall hit rates for effective and ineffective work behaviors
appear in Table 2.

**Effective Work Behavior Rating.** Analysis of variance revealed only a
main effect for the Individual vs. Group manipulation, \( F(1,88) = 23.32,\)
\( p < .0001, \) indicating that despite having read the same vignette,
groups reaching collective decisions attributed more effective behavior
to the police officer than did the individual subjects, mean hit rates for
groups and individuals equal to .92 and .81 respectively. The second
question addressed in this research is why? Did groups simply have a
memory advantage over individuals, or did they adopt a too-liberal
decision criterion? Measures of memory sensitivity (Pr) and response
bias (Br) were employed to determine whether either, or both, of
these mediating processes were responsible.

**Pr** is a widely used measure of memory sensitivity and provides
an index of information actually stored in memory. It ranges from -1.0
indicating no memory to 1.0 indicating perfect memory and is
computed as follows: \(^2\)

\[ Pr = \text{Hit Rate} - \text{False Alarm Rate} \]

**Br** is a recently developed measure of response bias. It ranges
from 0.0 to 1.0 with .50 indicating a neutral decision criterion (e.g., no
bias). Values of less than 0.5 indicate a too-conservative decision
criterion (e.g. bias to say a particular behavior did not occur), whereas
values greater than 0.5 indicate a too-liberal decision criterion (e.g.
<table>
<thead>
<tr>
<th>Groups</th>
<th>Individuals</th>
<th>Delays</th>
<th>Groups</th>
<th>Individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>48</td>
<td>18</td>
<td>44</td>
<td>28</td>
</tr>
<tr>
<td>72</td>
<td>62</td>
<td>10</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>68</td>
<td>55</td>
<td>0.21</td>
<td>86</td>
<td>0.72</td>
</tr>
<tr>
<td>63</td>
<td>52</td>
<td>4.64</td>
<td>52</td>
<td>4.55</td>
</tr>
<tr>
<td>16</td>
<td>7.0</td>
<td>52.1</td>
<td>7.0</td>
<td>9.69</td>
</tr>
<tr>
<td>9.1</td>
<td>22</td>
<td>4.78</td>
<td>27</td>
<td>8.78</td>
</tr>
<tr>
<td>11</td>
<td>15</td>
<td>0.98</td>
<td>8.3</td>
<td>8.42</td>
</tr>
<tr>
<td>22</td>
<td>19</td>
<td>0.91</td>
<td>5.9</td>
<td>5.92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effective Work Behavior</th>
</tr>
</thead>
</table>
| Mean Hit Rate, False Alarm Rate, Memory Sensitivity (P), and Response Bias (P)

Table 2
bias to say a particular behavior did occur). \( Br \) is computed as follows:

\[
Br = \frac{False \text{ Alarm Rate}}{1-(Hit \text{ Rate} - False \text{ Alarm Rate})}
\]

It is worth noting that this set of measures, unlike others commonly used, are independent and thus are ideal for pinpointing mediating processes (see Snodgrass & Corwin, 1988). To calculate \( Pr \) and \( Br \), hit rates and false alarm rates were transformed into measures of memory sensitivity \( (Pr) \) and response bias \( (Br) \) (see Table 2).

Preliminary analysis indicated that \( Pr \) and \( Br \) were uncorrelated \((r = .12, \text{n.s.})\).

**Memory Sensitivity \((Pr)\).** Analysis of variance produced a main effect for Time of Rating, \( F(1, 88) = 4.90, p < .03 \). More importantly, a main effect for Subject Type, \( F(1, 88) = 11.76, p < .0009 \), indicated that, even after correcting for guessing, groups were indeed more accurate than individuals in their recognition of previously displayed behaviors, \( Pr \) (memory sensitivity) means equal to .71 and .58 respectively.

**Response Bias \((Br)\)** A main effect for Subject Type, \( F(1, 88) = 6.73, p < .01 \), revealed that groups adopted a significantly more liberal decision criterion than did individual subjects (whose decision criterion was close to neutral). Means were equal to .67 and .53 respectively.

The above Anova results indicate that the higher attribution of effective behaviors (hit rate) by groups was a function of both memory sensitivity \( (Pr) \) and a systematic response bias \( (Br) \). Groups did indeed enjoy a memory advantage over individuals (even when correcting for guessing), but they also adopted a too-liberal decision criterion as
compared to individuals.

**Ineffective Work Behavior Rating.** Analysis of variance produced main
effects for Subject Type, $F(1,88) = 15.13$, $p < .0002$, and Time of
Rating, $F(1,88) = 8.06$, $p < .005$, and a significant interaction $F(1,88)
= 4.78$, $p < .03$. Cell-wise comparisons revealed that in the delayed
rating condition, groups attributed more ineffective work behaviors to
the police officer than did the individual subjects, mean hit rates equal
to .88 and .71 ($t = 3.31$, $p < .002$) respectively. No differences
between groups and individuals were found in the immediate rating
condition, mean hit rates equal to .89 and .85 ($t = 1.61$, n.s.). Again,
the important issue becomes whether these differences occurred
because of a memory advantage, a response bias, or a combination of
both. The independence of the $Pr$ and $Br$ measures was once again
confirmed, $r$ equal to -.13, n.s.

**Memory Sensitivity (Pr).** Analysis of variance produced a main
effect for Subject Type, $F(1,88) = 13.10$, $p < .001$, a main effect for
Time of Rating, $F(1,88) = 3.58$, $p < .01$, and a significant interaction,
$F(1,88) = 5.84$, $p < .01$. Follow-up tests revealed that in the delayed
rating condition groups were more accurate in remembering work
behavior than were individuals, whereas there was no significant
difference in the immediate rating condition. The mean $Pr$ scores for
groups and individuals in the delayed condition were .72 and .44, ($t =
3.72$, $p < .001$) respectively, and for the immediate condition, they
were .70 and .64 ($t = 0.97$, n.s.), respectively. It appears, then, that
after a five day's delay, memory sensitivity deteriorated in the same
individual condition ($t = 3.11$, $p < .003$), but was maintained at the
level in the group condition ($t = .43, \text{n.s.}$).

**Response Bias (Br)** Analysis of variance did not reveal any significant main effects or interactions ($p's > .30$).

**Relationship between Evaluative Judgment and Decision Criterion**

Pearson correlations between the mean Work Performance Score and the Response Bias (Br) measure for both effective and ineffective behaviors were computed for different experimental conditions (see Table 3).

As hypothesized, the overall (across subject and time conditions) evaluative rating correlated positively with Response Bias for the rating of effective behaviors ($r = .31, p < .003$) and correlated negatively for the rating of ineffective behaviors ($r = -.25, p < .02$). Further analysis revealed that this relationship was only significant for individual raters. The correlation between individuals' evaluative rating and their decision criterion when rating effective behaviors was $.38 (p < .004)$, and when rating ineffective behaviors, $r = -.30 (p < .03)$. For groups, on the other hand, the corresponding correlations were $.17 \text{ (n.s.)}$ and $-.18 \text{ (n. s.)}$. It appears then, that individuals' decision criterion was significantly related to their previous evaluative judgment, while that of groups was not.

**Confidence Level Analysis**

Two different approaches were used to analyze confidence levels. In the first analysis, the question was: are groups more likely to choose the option "we are positive that the behavior did (did not) occur," than individual subjects? Although a somewhat indirect measure of confidence, this approach yields insight into differences
Table 3

Simple Correlations Between Evaluative Rating and Response Bias for Effective and Ineffective Behaviors

<table>
<thead>
<tr>
<th></th>
<th>Effective Behaviors</th>
<th>Ineffective Behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>All subjects (n=89)</td>
<td>.31 (ρ &lt; .003)</td>
<td>-.25 (ρ &lt; .02)</td>
</tr>
<tr>
<td>Individuals (n=55)</td>
<td>.38 (ρ &lt; .004)</td>
<td>-.30 (ρ &lt; .03)</td>
</tr>
<tr>
<td>Groups (n=34)</td>
<td>.17 (ρ &lt; .34)</td>
<td>-.18 (ρ &lt; .30)</td>
</tr>
</tbody>
</table>
between experimental groups in terms of how subjects felt about their responses. The number of the "positive" responses (out of 32 responses) was calculated for each group and for each individual subject and the differences were analyzed in a 2 x 2 (Subject Type: Individual/Group x Time of Rating: immediate/delayed) between group analysis of variance. The average number of these type of responses for the different conditions are listed in Table 4.

Analysis of variance produced main effects for Type of Subject, \( F(1,88) = 44.37, \ p < .0001 \), Time of Rating, \( F(1,88) = 12.44, \ p < .0007 \) and a significant interaction, \( F(1,88) = 18.09, \ p < .0001 \). Follow up tests revealed that in the delayed rating condition groups chose the "we are positive" response more often than individuals (\( t = 9.97, \ p < .0001 \)), means equal to 24 and 11 times respectively. Differences between groups and individuals in the immediate condition were only marginally significant (\( t = 1.98, \ p < .06 \)), means equal to 24 and 21 times respectively. Groups said that they were "positive" of their response 75% of the time, and individuals only 48% of the time.

Furthermore, groups chose this response in the delayed condition as often as they did in the immediate condition. Individuals, on the other hand, indicated they were "positive" 62% of the time in the immediate condition and only 34% of the time in the delayed rating condition. Our hypotheses were therefore partially supported; groups indicated more certainty than individuals, but only in the delayed condition and delay of five days affected only the confidence level of individual subjects.

The second approach to the evaluation of confidence was an
Table 4

Mean Number of "Positive" Responses$^{1,2}$
For Different Experimental Conditions

<table>
<thead>
<tr>
<th>Subject Condition</th>
<th>Immediate</th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individuals</td>
<td>21</td>
<td>11</td>
</tr>
<tr>
<td>(sd= 6.6)</td>
<td>(sd=4.3)</td>
<td></td>
</tr>
<tr>
<td>n=32</td>
<td>n=23</td>
<td></td>
</tr>
<tr>
<td>Groups</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>(sd=4.5)</td>
<td>(sd=3.6)</td>
<td></td>
</tr>
<tr>
<td>n=19</td>
<td>n=15</td>
<td></td>
</tr>
</tbody>
</table>

$^1$ A "positive" response is defined as a response of, "I(we) am(are) positive that he did(did not)" on the behavioral rating task.

$^2$ Out of 32 responses on the Behavioral Rating Task.
attempt at a more precise measure of confidence that would
differentiate the confidence levels for correct and incorrect
responses. Correct responses were either "hits" or "correct
rejections," that is, subjects were considered to be correct when they
correctly identified "old" or "new" behaviors. "False alarms" and
"misses" were counted as incorrect responses. Furthermore, for both
correct and incorrect responses, ratings were transformed such that
"certain", "fairly certain" and "undecided" options were now given a
score of 3, 2 and 1 respectively. Total scores for correct and for
incorrect responses were computed for each group and for each
individual subject and these scores then constituted the confidence
score with 3 as the highest level of confidence. The mean confidence
level across all conditions was 2.23. Mean confidence scores for the
different conditions appear in Table 5.

A 2 x 2 x 2 mixed factorial ANOVA was used to test the effect of
Subject Type, Time of Rating, Type of Response (within-subjects
factor, correct vs. incorrect) and their interaction on the confidence
score. There was a main effect of Type of Response, (F(1,78) = 101.5,
p < .0001) with mean confidence levels for correct responses being
2.53 and that for incorrect responses, 1.94, demonstrating that
confidence was significantly higher for correct responses. Also, as
shown by an interaction of Type of Response and Type of Subject
(F(1,78) = 4.19 p < .004), this difference was larger for groups than
for individual subjects. Follow-up analyses showed no significant
simple effects or interactions for incorrect responses (p's > .1). For
correct responses however, there were simple effects of Type of
Table 5

Mean Confidence Scores For Correct and Incorrect Responses in Different Experimental Conditions

<table>
<thead>
<tr>
<th>Time of Rating</th>
<th>Correct Responses</th>
<th>Incorrect Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Immediate</td>
<td>Delayed</td>
</tr>
<tr>
<td>Individuals</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.56 (sd=.29)</td>
<td>2.17 (sd=.20)</td>
</tr>
<tr>
<td>Groups</td>
<td>2.65 (sd=.29)</td>
<td>2.69 (sd=.15)</td>
</tr>
</tbody>
</table>

Confidence Level Across all Conditions

2.23 (sd=.41)

Range: 3.0 = "positive"
2.0 = "fairly certain"
1.0 = "undecided"
Subject ($F(1,78) = 28.15, p < .0001$), Time of Rating ($F(1,78) = 9.20, p < .003$), and an interaction effect ($F(1,78) = 12.71, p < .001$). Cell-wise comparison tests revealed that for correct responses, group and individual confidence levels differed in the delayed rating condition ($t = 8.97, p < .0001$), means equal to 2.69 and 2.17 respectively, but not in the immediate rating condition ($t = .97$ n.s.), means equal to 2.65 and 2.56 respectively.

This second analysis elaborates on, and supports the first. Here it was shown that groups do indeed respond more confidently than individuals. However, this was evident only when groups responded correctly in the delayed rating condition. An additional, and maybe not very surprising finding, is that confidence levels were generally higher in the immediate condition.

Discussion

This was the first study to examine group vs. individual differences in terms of the two critical components of behavioral rating accuracy. The objective was to determine whether groups are more or less accurate than individuals in their memory of work behaviors and/or whether they adopt a more or less liberal decision criterion. A time of rating factor was included to determine whether a five day interval between stimulus presentation and measurement would moderate the differences between groups and individuals. Another objective was to explore differences in leniency or severity between groups and individuals on evaluative ratings. Lastly, it was hypothesized that evaluative ratings would be positively related to response bias when subjects rate the occurrence of effective behaviors.
and negatively related to response bias when subjects rate the occurrence of ineffective behaviors.

**Evaluative Ratings**

Groups evaluative ratings were not different than those made by individual subjects. However, it would be premature to conclude, based on this finding, that collective evaluative judgment is not qualitatively different than that of a subject working alone. One has to keep in mind that in this study only one target was evaluated. Thus, there was an absence of variability in terms of the multiple levels of performance that would normally be evident when several individuals are being evaluated. This may well have restricted the probability of detecting any existing differences in evaluative judgment among groups and individuals.

**Memory Sensitivity**

Without doubt, previous statements on the superior memory capacity of groups have been unequivocally upheld. Memory sensitivity for both effective and ineffective behaviors was much higher in groups than in individual subjects. The obvious question now is why? Why do groups have this superior ability of groups to be more discriminating in their memory of behavioral information than individuals?

For the rating task used in this study, "four heads" may, in fact, have been "better than one." With a group of interacting individuals faced with a recognition memory task, there is likely to be "pooling" or sharing of information. The aggregation of various pieces of information continues until the correct decision becomes obvious to all, or most members. Group members may have actively reminded
each other of how the police officer had behaved in the vignette and because of "cues" received during active discussion of the stimulus material, they were able to retrieve the necessary information. Thus the group, making use of its aggregate resources, correctly remembered more information than did the individual subject. Subjects in the individual condition had no opportunity for discussion and therefore received no cues to supplement their own retrieval process. A recording and analysis of group discussions may have provided at least partial support for this "cuing" explanation of superior group memory. Future efforts directed at identifying underlying factors of group vs. individual memory should therefore also include the collection and analysis of group process data.

Alternatively, the superior performance by groups in this study could have been due to the greater likelihood of there being an individual in a group who has above average memory ability. This reasoning is based on several models that have been proposed to account for the observed superiority of groups in memory tasks (Lorge & Solomon's Model A. 1955; Taylor, 1954; Thomas & Fink's Rational Model, 1961; see also Criswell, J., Solomon, H., & Suppes, P. 1962). Two key process assumptions of Lorge & Solomon's well-researched model are: 1) if one or more members know the correct answer, they will tell the rest of the group and 2) the group will always recognize this response as being the correct answer. The second assumption depends in turn, on several different structural and process group variables, such as the perceived status, verbosity, aggressiveness, similarity to other members, of the member making the suggestion. It
would be extremely difficult to measure, let alone control, such factors. Furthermore, since the task used in this study was essentially a recognition, rather than a recall task, this model is somewhat awkward to apply and because individual group member data was not collected, impossible to confirm.

Blau and Scott, (1962) suggested that one reason for the observed superiority of groups over individual performance in problem solving is the error-catching component of social interaction. In this study, the "cuing" that is assumed to have taken place during group discussion could have also served an error-catching function. Here again, recordings of group discussion would have indicated the extent of overt "corrections" that occurred. The chances that erroneous suggestions or responses are detected are understandably much higher in a group situation. However, the assumption that groups have a capacity to limit errors must be made with some caution. It may very well depend on the type of error. In this study, it was also shown that groups adopted a too-liberal decision criterion than individuals when judging the previous occurrence of behaviors. The adoption of a "too-liberal" or a "too-conservative" decision criterion is an erroneous strategy and if group interaction did indeed limit all types of errors, then the collective decision criterion should have been closer to a neutral one.

Response Bias

The findings confirm that there is a memory superiority associated with collective performance on a behavioral rating task. There is, however, the important qualification that this superiority
only partially accounts for the observed differences between group and individual performance on behavioral rating tasks. When unsure, groups tend to adopt a too-liberal decision criterion when rating the occurrence of effective behaviors. The differences in decision criterion demonstrated here indicate that one cannot discuss behavioral rating accuracy without a consideration of response bias.

As mentioned earlier, the propensity for bias in groups could have been the result of perceived social pressure on the part of individual group members. When a group of interacting individuals is attempting to reach consensus, group members often feel that they have to "give in" to what seems to be the predominant view. So in effect, the decision process becomes one of "majority rule."

Additionally, it may have felt "safer" or more socially acceptable for individual members to say that an effective behavior had occurred than to attempt to provide evidence that it did not. Consequently, there may have already been a slight individual bias to say that positive behaviors had occurred (in fact, a slight bias was evident, see Table 2). The fact that groups demonstrated a too-liberal decision criterion only when rating effective behaviors supports the possibility of the "polarization" of a slightly biased decision criterion that may have already been present in individuals.

The issue of group propensity for response bias is by no means clear. Research efforts should be directed at the factors that moderate this bias. Again, it would be profitable to measure how decision criterion changes when an individual joins a decision-making group. This could be accomplished by collecting individual group member
data. Also, the use of group process measures could possibly identify the social "forces" or interaction patterns that may be related to liberal decision-making.

Evaluative Judgment and Response Bias

The hypothesis that prior evaluative judgments would be positively related to decision criterion for the rating of effective behaviors and negatively related to decision criterion for the rating of ineffective behaviors was supported only in the case of individual subjects. The more favorable these subjects' evaluation of the police officer's performance before the behavioral rating task, the more likely they were to adopt a more liberal decision criterion when rating effective behaviors and a more conservative decision criterion when rating ineffective behaviors. Unfortunately, since the direction of the subjects' evaluations was a naturally occurring, rather than a manipulated factor in this study, and since only correlational data were available, no causal attributions can be made in terms of the effect of previous evaluative decisions on behavioral rating. Recently, Martell & Guzzo (in press), and Martell & Willis (1991) have shown that subjects who had previously been provided with positive performance cues adopted a more liberal decision criterion when judging the occurrence of effective work behaviors. When rating the occurrence of ineffective work performance, the same subjects adopted a more conservative decision criterion. While restating the critical qualification that no causal assumptions can be made based on correlational data, the pattern of the relationship between individual subjects' prior judgments and response bias found in the present study
directly parallels the findings relating performance cues and response bias. Whether subjects' own prior judgments can indeed act as "internal performance cues" is an empirical question that calls for more thorough experimental investigation.

Another point of interest here is the fact that this relationship between prior evaluative judgment and response bias only held for subjects working alone. No one explanation offers itself here. If prior judgments indeed acted as some sort of performance cues, why did these not also affect collective behavior on the rating task? It is possible that this effect is somehow "diluted" after group interaction. Alternatively, individuals may not have felt enough "ownership" of the group evaluative judgment and may therefore not have been as "psychologically committed" to it as were subjects who worked alone. In this case, since the prior judgment was not as "salient," it may not rigorous experimentation with these factors is needed before any have served as an "internal performance cue." In any case, more definite hypotheses can be formulated.

Levels of Confidence

In the first analysis, the level of confidence was operationalized as the number of times subjects chose the response that indicated the most certainty. The difference between groups and individuals varied as a function of the time of measurement. While there was only a marginal difference between groups and individuals in the immediate condition, this effect was much larger after a delay of five days, with groups choosing the "we are positive" response twice as frequently as individual subjects. The second analysis elaborated on the first in two
ways. First, level of confidence was now indicated by the score on a 3-point scale (transformation of the 6-point rating scale) with 1 indicating an "undecided" response and 3, a "positive" response. Second, confidence level was examined in terms of whether the response was correct and incorrect. The findings indicate that differences in confidence levels between the two types of subjects and the two time of rating conditions were evident only for correct responses. Here, group confidence level was higher than that for individual subjects and groups were equally confident in both time of rating conditions while individual subjects were less confident in the delayed condition.

The most intuitively appealing explanation for these findings is that groups were more certain of their answers because of the "affirmation" received during group discussion when more than two subjects agreed on a response. In other words, group members were confident of the consensus reached by the group because they may have felt that if the majority of the group agreed on whether a behavior had occurred or not, then it was probably right. The fact that confidence was much lower for those group responses that were ultimately incorrect suggests that these may have involved more disagreement which then reduced the "certainty" of the consensus decision. (Again, this could be confirmed only by attention to process variables such as amount of discussion and level of disagreement). Individual subjects received no outside support or "affirmation" for their responses and therefore may have generally been less sure of their responses. Their confidence level was much lower in the
delayed condition, presumably because of a deterioration in memory. For groups, the "cueing" that is thought to have occurred during group discussion may have sustained not only their memory, but also the confidence in their responses.

The Stephenson et al. (1986) findings mentioned earlier, namely that groups' confidence in their incorrect responses was "outrageously high" were not supported in the present study. It has to be kept in mind, however, that the task used in that study was one of recall, and not recognition. Also, unlike in the present research, confidence levels were obtained by directly asking subjects to indicate their confidence in their responses. For these reasons, findings of the Stephenson et al. study and findings from the present study may not be directly comparable.

Lastly, the concept of liberal decision strategy seems to be related to the issue of confidence. Given that one is not confident in his/her own judgment, one is more likely to be cautious or conservative when making decisions. If one is very confident, on the other hand, he/she may have a tendency to be liberal when responding to a task such as the one included in this study. Groups may have overestimated their potential for accuracy - maybe because of the conventional wisdom about "four heads" being better than one; they may therefore have processed information at a "shallower" level and reached agreement before all response options had been thoroughly examined. However, more experimental control would be needed before a link between confidence level and response bias can be clearly established. For instance, level of confidence could be included
as an experimental factor and response bias of subjects at different confidence levels can be compared. Classification into groups could be based on pretesting general confidence levels or (probably more efficiently) on manipulation checks after subjects have been given false performance feedback aimed at inducing high or low levels of task-specific confidence.

Practical Ramifications

As mentioned earlier, there are important practical implications of behavioral rating accuracy for performance appraisal, especially that which involves the use of behavioral checklists. The finding that group memory was more accurate, and that they attributed more behaviors correctly than individuals is quite encouraging. It suggests that random errors in rating accuracy may be controlled by investing the energies of a group of individuals and "pooling resources" for the task of appraising performance. When it comes to more systematic types of errors, however, group judgment may hurt the performance appraisal process. In this study, groups rating rating effective behaviors demonstrated a liberal response bias. In a realistic work setting, this may well result in "undeservingly" favorable performance appraisals. A performance appraisal panel may therefore not be effective in identifying the areas in which an employee needs further development. The "group memory advantage" then, has to be weighed against the inevitable loss of potential accuracy that will result from the groups' tendency to be liberal in their responses. In addition, future research could determine whether group judgment is also that susceptible to other forms of systematic errors or biases, such as those
result from stereotyping or the influence of initial expectations.

It seems that the demonstrated superiority of groups on this type of task should be taken advantage of but there is also the need to "de-bias" individuals who will be involved in collective performance appraisal. Martell & Willis (1991) point out that attempts to reduce bias and distortion by improving appraisers' memory (such as through the use of behavioral diaries or observational training) may be misdirected. Interventions may be more profitably directed at changing, or neutralizing, appraisers' decision criterion when they are retrieving behavioral information. These authors suggest that simply making raters more aware of their decision criterion when judging the occurrence of certain behaviors may, in fact, interrupt an otherwise automatic process that appraisers are not aware of. If performance appraisal groups could acquire a meta-cognitive skill of monitoring their own decision strategies, accuracy of their behavioral rating may more closely approach their "potential" accuracy.

This research has examined behavioral rating accuracy in terms of a model of recognition memory that differentiates between memory sensitivity and response bias. It should be noted, however, that this is by no means the only framework in which rating accuracy can be examined. In fact, a thorough understanding of group vs. individual differences in performance appraisal accuracy can only come about through the convergence of findings from studies that use different conceptualizations of accuracy. Other accuracy measures that have been used in the examination of performance ratings, are Cronbach's accuracy scores (see Becker & Cardy, 1986 and Murphy, Balzer,
Kelm, & Armstrong, 1984), distance accuracy (see McIntyre, Smith & Hassett, 1984 and Heneman & Wexley, 1983), and Borman's (1977) differential accuracy (see Cardy & Kehoe, 1984 & Pulakos, 1984). The use of more than one operationalization of rating accuracy is important since as Sulsky and Balzer (1988) point out, different measures may tap different facets of rating ability. Thus, the present study represents only a preliminary analysis of group vs. individual differences in behavioral rating performance.

In assessing the generalizability of the present findings, one has to consider two external validity issues that have repeatedly been brought up in relation to laboratory research in performance appraisal. The first of these is the "paper people" problem. In the present study, the police officer being rated was not a "real" individual but a "paper person." Despite all the research dedicated to this issue, there does not seem to be compelling differences among findings of laboratory and that of field research (see Dipboye, 1990, for a review). Murphy, Herr, Lockhart & Maguire (1986), in a comparative meta-analysis, contrasted the outcomes of paper people studies to those of similar studies in which ratings were based on the actual observation of ratee behavior. They offered a "signal-to-noise ratio" hypothesis as an explanation of why effect sizes may be larger in paper person studies. It was suggested that in this type of study, the performance-relevant "signal" is much stronger than when behavior is directly observed and that there is relatively little background noise (performance-irrelevant information). What was found, in fact, was that among the five different categories of studies included in the meta-analysis,
effect sizes for "paper people" studies were larger only in those studies that manipulated purpose of rating. Furthermore, there were only two studies in this category, so the authors themselves urged caution in the interpretation of the findings. In the present research, the "signal-to-noise ratio" hypothesis would have predicted that the "signal", in this case measured in terms of memory sensitivity ($Pr$), would be close to maximum since the ratee was a paper person. In fact, $Pr$ values were not near the ceiling at all (overall $Pr = .63$), so it is unlikely that the effect size in this study was a function of the stimulus presentation mode. There is additional evidence supporting the generalizability of "paper people" research. A study by Dobbins, Cardy and Truxillo (1988) did not find differences in results from a laboratory study involving written vignettes and a field study involving the actual observation of behaviors. Woehr and Lance (in press) found no difference in behavioral accuracy and decision criteria for subjects viewing ratee behavior on a videotape and those reading about it in a script. Furthermore, these authors warn that the paper person versus direct observation distinction may be too simplistic and upheld Dipboye's (1985) recommendations for laboratory research to continue to theoretically simulate the relevant aspects of on-the-job settings.

The second external validity issue pertains to the use of college students as raters. Numerous studies have compared student and nonstudent performance (e.g. Olian, Schwab & Haberfeld, 1988; Mullins, 1982; McGovern, Jones, & Morris, 1979; Bernstein, Hakel, & Harlan, 1975; Hakel, Dobmeyer & Dunnette, 1970). Yet, there is very little evidence that students are inappropriate subjects. The
general conclusion seems to be that when, as in the present research, the focus is on judgmental processes, no significant differences have been found between student and non-student performance (Dipboye, 1985). In the case of this study which was aimed at discovering differences between group and individual judgment, this issue may be of even lesser importance. There is no doubt of the need to study the performance of "real" groups rating "real" people in "real" settings. It is likely, however, that the group processes facilitating behavioral rating are present regardless of what type of subjects make up the group.

It is hoped that the issues raised by this study will help to develop a theoretical framework for the understanding of group vs. individual differences in information processing that goes beyond the "laboratory versus field" distinction. This research has extended and reinforced previous findings of group memory superiority and, more importantly, it has provided strong evidence of a propensity for response bias in group behavioral judgment. The stage is now set for more intensive investigation of the factors responsible for differences in memory and response bias between groups and individuals.
References


Cronbach, L. J. (1955). Processes affecting scores on "understanding
of others" and "assumed similarity". Psychological Bulletin, 52, 177-193.


Organizational Behavior and Human Decision Processes, 41, 180-195.


and Social Psychology, 50, 1113-1122.
Footnotes

1. These measures are explained in more detail in the Results section.

2. Note that the hit rate is the conditional probability of responding "yes" to a previously observed behavior, \[\text{HR} = p("yes"/"old" \text{ behavior})\]. The false alarm rate is the conditional probability of responding "yes" to a previously unobserved behavior, \[\text{FAR} = p("yes"/"new" \text{ behavior})\]. As is recommended (Snodgrass & Corwin, 1988; Upton, 1987), the hit and false alarm rates were corrected prior to analysis by adding 0.5 to each frequency and then dividing by \(N + 1\) (\(N\) = number of old or new items).
Appendix A

Vignette
JOB DESCRIPTION: POLICE OFFICER

The job of a police officer is to assist in maintaining peace and public order. To do this, a police officer has three major job responsibilities. First, a police officer must act to prevent violations of the law. Second, a police officer must observe for any possible violations of the law. Third, when a law is broken, a police officer is responsible for its enforcement. This requires police officers to be alert, level headed, physically fit and able to enforce police procedures in a variety of situations, some of which may be unusual or potentially dangerous.

BIOGRAPHICAL INFORMATION

NAME: John Dutton
AGE: 22 years old
BIRTHPLACE: Born and raised in a suburb outside Boston, Mass.
HOBBIES: Enjoys bowling and going to the movies
PHYSICAL DESCRIPTION: 6' 0"; 185 lbs.
OCCUPATION: Police Officer
             John Dutton is with the Boston Police department.

Now that you have a general impression of the job and of the job-holder, please turn to the next page and begin reading the vignette.
WORK VIGNETTE

DAY ONE

It was 8:48 am and Officer John Dutton was chatting with several fellow officers who were just coming off their shifts. Several officers were laughing and kidding another officer who had accidently suffered a minor burn by kicking a lighted flare off the roadway. Since there was twenty minutes before "detail" was to begin, Officer Dutton decided to read last month's issue of "Police Monthly." He found that there was a great deal of valuable information concerning court rulings and pending legislation on police-related topics such as arrest procedure rulings, interrogation procedures, and Miranda rulings. Officer Dutton was surprised that other officers did not spend more time keeping up on this type of police-related information.

After "detail," Officer Dutton walked to his assigned squad car wondering what today would bring. Before driving off, he took a minute to look under the seats to be sure that no drugs or weapons had been left behind by anyone transported on the previous shift. He then immediately drove to his assigned beat.

The first call of the shift directed him to a nearby shopping mall. The dispatcher informed him that there was a "pursuit-in-progress" and that he should set-up in the parking lot of the mall and await further instructions. While driving to the location, Officer Dutton noticed that the car was responding poorly. The steering wheel vibrated and the brakes were soft. Since he had not examined the car before
leaving and because it was too late to get another car, Officer Dutton continued on. Once in place, the dispatcher instructed Officer Dutton to watch for a 1987 blue Ford Torino driven by the suspect. After several minutes, Officer Dutton realized that he was having trouble identifying different types of cars. Not wanting to chance missing the suspect's car, Officer Dutton called in to the dispatcher to ask for a description of this type of car. He was informed that the chase had ended in another part of town and that he could return into service.

Back on the beat, Officer Dutton continued on patrol. After driving around for an hour or so, he decided to patrol on foot. Officer Dutton found this to be a good way of meeting people and maintaining a presence. During the last few weeks he had introduced himself to some of the more colorful street people on his beat in an attempt to get to know them. As he turned the corner Officer Dutton saw two such street people and talked with them for a while. They talked about a recent rash of purse snatchings and who could be responsible. Just then, a citizen's complaint about a floating crap game came in over the radio.

Responding to the complaint, Officer Dutton did indeed find a crap game several blocks away. Officer Dutton approached the four men involved in the illegal game. As they dispersed, a pocket knife dropped out of the pocket of one of the players. Officer Dutton recognized the man as a parolee just out of prison on an early release program. After the others had left, Officer Dutton took the man aside
and informed him that he was in violation of his parole. Officer Dutton told the man that if he would serve as his eyes and ears on the street that he would drop the matter. Officer Dutton told the man that he had one week to think it over. The remainder of the shift was pretty calm with the usual number of minor calls and disturbances.

**DAY TWO**

The next day, Officer Dutton began his shift by taking a ride over to the local pool hall to talk to some of the neighborhood "toughs" about staying out of trouble and bettering themselves. He drove the squad car up to where a bunch of kids were hanging around. After introducing himself he talked and hung out for about 20 minutes.

Back on the beat, Officer Dutton pulled over a car for failing to stop at a red light. He told the dispatcher that he was leaving service to issue a traffic ticket and approached the car. Once at the car, he asked the driver for his license and registration. He informed the driver that he had gone through a red light and then issued the ticket. The driver accepted the ticket with little comment and drove off. Back in the squad car, Officer Dutton finished the traffic violation report and then informed the dispatcher that he was back in service. "That's fine" said the dispatcher," but you forgot to call in the plate number to confirm the car ownership." The dispatcher told him not to worry about it and to just continue his tour of duty.
Suddenly a call came in over the radio. It was not directed at him, but since he was in the area Officer Dutton decided to answer the call anyway. He radioed the dispatcher that he was in the area and would be happy to provide back-up. Officer Dutton's job was to use his squad car to block off a possible escape route. Officer Dutton did as he was asked and waited. After 10 minutes a call came in that the suspects had been captured.

Officer Dutton's next stop was the court house where he was needed to collaborate the testimony of several citizens who had seen a young man leaving an apartment building with a stolen television set under his arm. The district attorney was deciding whether there was sufficient evidence to charge the man and thought that Officer Dutton's testimony, although not crucial, might be helpful. As it turned out, Officer Dutton was not very helpful. He had difficulty remembering the exact time of the day the alleged crime took place, nor could he be absolutely certain of many of the details. Officer Dutton complained that the incident had happened almost two weeks ago and he could not be expected to remember anything.

Back on the beat, Officer Dutton pulled a car over for speeding through a busy intersection. The driver, a middle-aged businessman, was very abusive and uncooperative. He argued that Officer Dutton had no right to pull him over. Officer Dutton did not get shook-up, he remained calm and in a steady, confident voice explained the violation and issued the driver a traffic citation.
Next, near the end of the evening shift, Officer Dutton drove through the business district. Since it was after business hours, Officer Dutton thought it was unusual when he saw a person hanging around in front of a used car lot that had closed several hours earlier. He got out of the squad car and began to question the man about who he was and then he radioed the dispatcher to see if there were any arrest warrants pending on the person. Then he told the man he was free to go, thanking him for his cooperation.

**DAY THREE**

The next day while on the beat Officer Dutton noticed a small crowd of people calling for help. As he approached, he saw a disheveled male about thirty years old lying on the ground. Officer Dutton examined the man but had no idea what could be wrong. He called in to the dispatcher for a medical unit and waited. While waiting, several people mentioned to him that the man's eyes looked funny and that maybe he should check the man's arms for needle marks, indicating a possible drug overdose. Officer Dutton checked the man's arms and sure enough found several fresh needle marks. He radioed the medical unit that this was probably a drug overdose.

Since it was now late evening and most of the stores had closed for the day, Officer Dutton continued to patrol the main street business district on foot. It was his habit to check the front and back doors of stores and other businesses for signs of possible break-ins and burglaries. While doing this, Officer Dutton heard a report of a minor
traffic accident nearby. He decided to ignore the call and assumed that other units would respond. After finishing his rounds, Officer Dutton climbed into his squad car and headed back toward the station house.

There were only 20 minutes left on his shift and Officer Dutton was looking forward to having the next two days off from work. While driving back to the station house, information regarding a teenager suspected of several burglaries on his beat came in over the radio. The suspect himself had not been seen in about a week and had apparently dropped out of sight. A good friend of the suspect had just been picked up and was now at the station house. Officer Dutton drove directly to the station to begin questioning the suspect's friend. He spent the next two hours talking with the individual. Since it was now pretty late the sergeant agreed to allow Officer Dutton to continue the questioning the following day.
Appendix B

Instructions
GENERAL INSTRUCTIONS

HI, I'm Mae Borg, a second year graduate student and I am working as a research assistant for Dr. Martell in the I/O Psychology program. We're interested in the performance appraisal process, specifically, "how is it that we evaluate another person's work behavior." In this study, we're most interested in your evaluations of a police officer. Before I tell you more about this study and what you will be doing, please read and sign the consent form, I will collect it when you're done.

Please open Booklet #1 and read to yourself the instructions that appear on the first page as I am reading them aloud to you.

"This is a study investigating how people perceive and evaluate an individual's work behavior. On the pages to follow, you will be reading a "vignette" that depicts the work behavior of a police officer. The vignette outlines the work behavior of an actual police officer over a period of 3 days. The vignette is about five pages long and is part of a police training guide currently in use. Also included is a job description and a brief biography of the police officer you will be reading about.

Your task is to read the vignette with the goal of evaluating the individual's work performance. You will be given approximately 15 minutes to read the vignette, then either today or in a few days, you will be asked to answer some questions about the police officer's work performance.

All your responses will be anonymous. Please do not put your name or any identification on any of the materials you are given. If at any time during the experiment, you have any questions, please feel free to raise your hand and the experimenter will try to answer your question." Are there any questions before we begin?
Remember you have approximately 15 minutes to read the vignette. When you are finished, please simply close the booklet and await further instructions. You may begin by turning to the next page.

Now that you've finished reading the vignette, please put away booklet #1 by placing it at the bottom of your stack of booklets. You may not refer back to this booklet throughout the rest of the experiment.

**DISTRACTOR TASK**

The first task we are asking you to do is a short one and is not directly related to what you have read. We are collecting some information from students about their opinions about certain issues concerning the development of this university. Please open BOOKLET #2, and read the instructions on the first page.

(2 minutes later)

If there are no questions, I will now start the clock, in five minutes I will ask you to stop working on this task. You may now turn the page and start.  
(In the Delayed Rating Condition, subjects are dismissed and given written instructions to return in five days to complete the experiment)

**EVALUATIVE TASK (INDIVIDUAL)**

Please stop working and put away BOOKLET #2 by placing it at the bottom of your stack of booklets. Open BOOKLET #3 and please read the instructions on the next page to yourself, as I am reading them........

"We are interested in your evaluation of the police officer John Dutton whom you read about in the vignette. Using the scale below, please circle the number which best describes your evaluation of the Police Officer. Please be sure to answer both items. You will have ten minutes for this task."

Are there any questions, if so please raise your hand.
If there are no questions, I will now start the clock, in ten minutes I will ask you to stop working. You may begin.

**BEHAVIORAL RATING TASK (INDIVIDUAL)**

Please stop working and close Booklet #4. Place this booklet at the bottom of the stack of materials. Now open Booklet #5 and carefully read the instructions on the first page.

(2 minutes later)

If there are no questions, I will now start the clock. You will have 30 minutes for this task. Please follow all instructions closely. You may begin.

**EVALUATIVE TASK (GROUP)**

Please stop working and put away BOOKLET #2 by placing it at the bottom of your stack of booklets. Open BOOKLET #3 and please read the instructions on the next page to yourself, as I am reading them.......

"We are interested in your evaluation of the police officer John Dutton whom you read about in the vignette. You will be doing this task AS A GROUP. You and the other three people at your table constitute one group. When I tell you to start, you will read the items below and DISCUSS WITH THE OTHER STUDENTS IN YOUR GROUP what rating to give each item. As a group, you will REACH CONSENSUS on the rating that is appropriate. What consensus means is that EVERYONE IN THE GROUP HAS TO AGREE on what rating is appropriate. When you have decided together on a rating, on group member should circle the number that best describes the group's evaluation of the Police Officer.

Before you start on the first item, you should decide quickly among yourselves who will be the scribe for the group, that is the one who will write down the ratings. The rest of the group members are not to mark in their booklets. There will be other tasks in the experiment
that will require you to make consensus decisions. The scribe for this task will continue to be the scribe for the rest of the experiment. You will have ten minutes to complete this task." Any questions so far, if so please raise your hand.

If there are no questions, I will now start the clock, in ten minutes I will ask you to stop working. You may begin.

**BEHAVIORAL RATING TASK (GROUP)**

Please stop working and close Booklet #4. Place this booklet at the bottom of the stack of materials. Now open Booklet #5 and carefully read the instructions on the first page.

(2 minutes later)

If there are no questions, I will now start the clock. You will have 30 minutes for this task. Please follow all instructions closely. You may begin.

Please stop working and put Booklet #5 at the bottom of your stack. You have now completed the experiment.

I would now like to tell you briefly about the experiment and try to answer any questions that you might have.

(DEBRIEFING FOLLOWS)
Appendix C

Distractor Task
In the event that Rice University were to receive a large endowment ($100,000) from a private source, and because of tax reasons, this money had to be used within 12 months; how should the money be allocated?

Please list up to six departments, services, or other aspects of the university that you think the money should be spent on. List these according to their priority to you as a student.

You will have 5 MINUTES to complete this task.

1. ____________________________________________

2. ____________________________________________

3. ____________________________________________

4. ____________________________________________

5. ____________________________________________

6. ____________________________________________

(STOP AND AWAIT FURTHER INSTRUCTIONS)
Appendix D

Evaluative Task Materials
Using the scales provided below, your next task is to evaluate the performance of police officer John Dutton. Please circle the number that best describes your evaluation of the police officer.

You will have **FIVE** minutes to complete this task.
If there are any questions, please raise your hand.

For the following questions, circle the number that best describes your evaluation of police officer John Dutton. It is critical to the success of this research that you answer all four questions. Please do not leave any questions blank.

1. How would you rate this individual's competence as a rookie police officer?
   
   Not Very 1 2 3 4 5 6 7 8 9 Very Competent

2. Overall, how would you rate this individual's job performance as a rookie police officer?
   
   Poor 1 2 3 4 5 6 7 8 9 Excellent

3. How would you rate this individual's potential for advancement as a police officer?
   
   Not Very Favorably 1 2 3 4 5 6 7 8 9 Very Favorably

4. How successful is this individual likely to be in the future as a police officer?
   
   Not Very Successful 1 2 3 4 5 6 7 8 9 Very Successful
Using the rating scales provided below, your next task is to evaluate the performance of police officer John Dutton. However, you will be evaluating the police officer AS A GROUP. You and the three other people with you constitute one group. Each group must agree together on their rating of the police officer. That is, you must reach consensus as a group concerning what rating is most appropriate. **EVERYONE IN THE GROUP MUST AGREE** on the rating. One member from each group has been randomly assigned as a SCRIBE for that group; their packet is clearly marked SCRIBE on the front. When you have decided together on a rating, the SCRIBE should circle the number that best describes the group's evaluation of the police officer in his or her booklet. The rest of the group members are not to mark in their booklets.

You will have **FIVE** minutes to complete this task. Any questions so far, if so please raise your hand.

For the following questions, circle the number that best describes the group's evaluation of police officer John Dutton. It is critical to the success of this research that the group answer all four questions. Please do not leave any questions blank.

1. How would your group rate this individual's competence as a rookie police officer?

   Not Very  1  2  3  4  5  6  7  8  9  Very Competent

2. Overall, how would your group rate this individual's job performance as a rookie police officer?

   Poor  1  2  3  4  5  6  7  8  9  Excellent

Turn to the next page
3. How would your group rate this individual's potential for advancement as a police officer?

Not Very Favorably
1 2 3 4 5 6 7 8 9 Very Favorably

4. How successful does your group think this individual is likely to be in the future as a police officer?

Not Very Successful
1 2 3 4 5 6 7 8 9 Very Successful
Appendix E

Behavioral Rating Task Materials
BEHAVIOR QUESTIONNAIRE

On the following pages you will find a list of specific work behaviors. Using the scale below, please rate whether the Police Officer, John Dutton performed the work behavior in the vignette you read.

The ratings are to be made by placing the number (1, 2, 3, 4, 5 or 6) which most closely represents your opinion, on the line next to each behavior.

1. I am positive that he did not.
2. I am fairly certain that he did not.
3. I am undecided, but think that he did not.
4. I am undecided, but think that he did.
5. I am fairly certain that he did.
6. I am positive that he did.

Please respond to each question. The success of this research requires that you answer every question. Please do not leave any blank. You have 25 minutes for this task.

You may begin by turning to the next page.
BEHAVIOR QUESTIONNAIRE

On the following pages you will find a list of specific work behaviors. Using the scale below, your group is to rate whether the Police Officer, John Dutton, performed the work behavior in the vignette you read. **ALL** group members must agree on the rating of each behavior. It is important that you agree together, as a group, on what rating to give each behavior; again, **ONLY** the SCRIBE for the group should actually write down the ratings.

The SCRIBE is to write the number (1, 2, 3, 4, 5 or 6) which most closely represents the group consensus, on the line next to each behavior.

1. We are positive that he **did not**.
2. We are fairly certain that he **did not**.
3. We are undecided, but think that he **did not**.
4. We are undecided, but think that he **did**.
5. We are fairly certain that he **did**.
6. We are positive that he **did**.

Please respond to each question. The success of this research requires that you answer every question. Please do not leave any blank. You have 25 minutes for this task.

You may begin by turning to the next page.
a. ____ checked daily bulletins concerning recent criminal activity, outstanding arrest warrants, pictures of suspects

b. ____ remained cool but firm when verbally abused by a citizen in the course of duty

c. ____ began shift without making sure that the patrol car was in working order

d. ____ worked overtime gathering information on a case

e. ____ shared information with other officers about unusual activity observed while on duty

f. ____ misunderstood a code assignment when responding to a call

g. ____ devoted extra time to patrolling "problem areas" of high criminal activity.

h. ____ handled a dispute between citizens by immediately separating the feuding parties

i. ____ before beginning shift, searched the squad car for any objects that may have been concealed by prisoners

j. ____ ignored a report of a nearby accident

k. ____ quickly determined whether an incident was a civil or a criminal matter

l. ____ directed other officers and citizens away from crime scene in order to protect important evidence

m. ____ deliberately covered beat in a different manner each day so as not to establish a set and predictable patrol pattern

n. ____ volunteered to help out on a call assigned to other officers
o. ___ acted confused and uncertain when testifying about a crime

p. ___ rattled doors while walking beat to check for signs of forced entry

q. ___ was inattentive during "detail"

r. ___ sought information about recent police-related court rulings and legislation

s. ___ checked to see if there were any outstanding warrants when questioning suspicious person

t. ___ behaved nervously when making an arrest

u. ___ forgot to check if a vehicle stopped for a traffic violation had been listed as stolen

v. ___ had difficulty locating street addresses when responding to a call

w. ___ stopped in a bar to check for any potential problems

x. ___ was slow to recognize a narcotics overdose

y. ___ discussed building security and crime prevention with citizens

z. ___ talked with local "toughs" about staying out of trouble

aa. ___ turned in an incomplete arrest report

bb. ___ informed citizens of recent criminal activity in the area
cc. ___ during a disturbance call, left partner alone in the house with the combatants

dd. ___ talked with known "characters" on the beat to gain valuable "street" information

ee. ___ had difficulty distinguishing among different models and years of cars

ff. ___ talked with a neighborhood ex-con about becoming a police informer

Please turn to the next page