Analyzing Vigilance Performance: Task-type, Feedback, and Predictors

Wendy A. Darr

A thesis submitted in Partial Fulfillment of the Requirement

For the Degree of Master of Science in

Applied Psychology (Industrial/Organizational)

Saint Mary's University

Halifax, Nova Scotia

Date of Submission: 30 September, 1999

Approved: ____________________
Dr. L.L. Methot (Thesis Advisor)

Approved: ____________________
Dr. V.M. Catano (Committee Member)

Approved: ____________________
Dr. K. Hill (Committee Member)

Approved: ____________________
Dr. J. N. Leary (Committee Member)

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

0-612-47677-4

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The present study was funded through the Saint Mary's University Senate Research Grant, presented to thesis advisor Dr. Laura L. Methot. I am especially grateful to Dr. Methot for involving me in this area of research three years ago, and for using her behavioural background in "shaping" my progress up to this point. The completion of this project would not have been realized without her ongoing feedback and direction. I am also grateful to Dr. Victor Catano and Dr. Kenneth Hill for their supervisory direction as members on my Thesis Advisory Committee (TAC). Special thanks to Dr. James Leary at the Department of Psychology, Dalhousie University for agreeing, at very short notice, to serve as a member on my TAC. Final notes of appreciation to Natalie Boltz for her assistance during the data-collection phase of this study and to my family and well-wishers for their supportive encouragement during the completion of this project.

APPENDICES

LIST OF TABLES

LIST OF FIGURES

Abstract

Two vigilance tasks, monitoring and inspection, were employed in the examination of the effects of feedback and predictors (personality and cognitive) on vigilance performance measured through hits, decrement, and false-alarm rate. The tasks differed on several characteristics including the type of discrimination, complexity, and operational relevance. A repeated-measures design was utilized in which 68 university students participated in computer simulations of both tasks. Differences in performance on the two tasks reveal that complexity may hinder hit performance, but may mitigate vigilance decrement and false alarm rate. Results also suggest that the usefulness of personality and cognitive predictors of vigilance performance may be task specific. Findings provide insights into the implementation of countermeasures such as introducing an element of complexity on simple tasks (perhaps through artificial signal injection) and reducing complexity through training on the inspection task. Insight into the inconsistencies surrounding previous attempts at predicting vigilance performance and into the delivery of feedback on vigilance tasks with complex visual displays is also provided.

Introduction

The Vigilance Phenomenon

Vigilance, the maintenance of accuracy in responding to critical changes in the environment over long periods of time is associated with tasks such as radar operations and assembly-line inspection (Davies & Parasuraman, 1982; Nachreiner,1977). Typically, individuals performing either of these tasks are required to monitor a display or some other stimulus in search of infrequent critical signals (Badalamente, 1969). Detailed examination of the vigilance phenomenon began in response to a serious applied military problem, leading N. H. Mackworth to experimentally determine the optimal length of watch for radar-operators. The Mackworth Clock test was designed such that performers watched for the deflection of a black pointer over a period of two hours. The Mackworth study revealed that performance on a simulated monitoring task rapidly deteriorated after about 30 minutes (Mackworth, 1948); this decline in performance came to be known as vigilance decrement (Davies & Parasurman, 1982).

Experimental vigilance tasks are designed to measure an individual's state of readiness to detect certain specified, random changes in the environment (J.F., Mackworth, 1970). However, tasks vary from one study to another because task specifications are determined partly by the factors one is interested in studying. A typical experimental paradigm involves the repeated presentation of infrequent signals (critical stimulus changes) together with background events or noise. The critical stimuli (signals), which are different from the background events on some dimension, require a response from the observer.

Measures typically used to determine the efficiency of human monitors on experimental vigilance tasks are overall hits (the number or percent of critical signals correctly detected), false alarms (the number of incorrect reports of a signal), and reaction time (latency in responding to a signal). Performance decrement is frequently calculated using either of these measures as a function of time. In other words, a decrease in hits over time or an increase in reaction time would be indicative of decrement in performance at a particular task. This decrement is assumed to reflect the performer's ability to sustain attention over time-on-task, also referred to as the performer's sensitivity to task stimuli (See, Howe, Warm & Dember, 1995).

Proponents of signal detection theory, however, point out that this estimate of performance may be biased as performers' responses may also be a result of non-perceptual factors (that is, performer's willingness to perform) rather than purely perceptual factors (that is, ability to perform). Signal detection theory has been employed in vigilance research to provide another measure of vigilance performance called sensitivity decrement. Sensitivity decrement, calculated as a function of target hits and false alarms, has been reported in several studies and it provides a bias-free estimate of the performer's actual perceptual sensitivity to the given task (for example, Parasuraman & Mouloua, 1987; Tanner, Jr. & Swets, 1954).

Goals of Vigilance Research

Research following the Mackworth studies was directed at either fitting *the task to the operator* or fitting *the operator to the task* (Craig, 1984). The former focussed on manipulating the vigilance task itself, either to reduce uncertainties pertaining to signals (for example, Johnston, Howell & Goldstein, 1966), to motivate the performer by making the task more interesting or challenging (for example, Baker 1960), and/or to optimize the surrounding conditions by including a moderate degree of stress (for example, Poulton, 1973). In essence, differences in performance across various vigilance tasks (for example, Buckner, Harabedian & McGrath, 1960) have essentially prompted researchers to look for task-specific features capable of maximizing vigilance performance.

On the other hand, efforts aimed at fitting *the operator to the task* came about in response to another consistent finding in vigilance research, interindividual differences in performance at the same vigilance task (Craig, 1984; Matthews & Holley, 1993). Such efforts prompted the search for predictors of vigilance performance, claiming that performance at such tasks can be improved by selecting individuals that have certain predispositions to be vigilant. Still other attempts at pin-pointing the source of inter-individual variation in performance suggest that such differences can be minimized by altering features of the task itself (Parasuraman, 1976; Methot & Huitema, 1998).

Vigilance Research: Some Criticisms

In pursuit of the attenuation of vigilance decrement, research expanding over the last four decades has focussed on either one or both of the above goals. Where has all this research brought us? How close have we come to solving the original vigilance problem since the days of Mackworth? Given four decades of research, one would assume that we have generated enough knowledge to help us tackle the inherent 'vigilance decrement' problem. Yet critics are far from optimistic. The main criticisms of vigilance research centre around two themes: there are few if any vigilance decrements in the real world, and even if an operational problem does exist, laboratory research being done (with relatively simple tasks) will generalize poorly to complex tasks in the operational setting (Adams, 1987).

In justification of their work many researchers disagree with Adams, claiming that vigilance failures, resulting from a decrement in performance or from low overall levels of vigilance, *can* be found in the real-world (Parasuraman, 1987). For example, Weiner (1987) cited a 1986 newspaper article, which described a new high-tech development in Automatic Train Operation (ATO) that allows the speed of the train to be controlled automatically, hence requiring minimal handling by train operators. However, train operators revealed that the system is far from perfect, adding that the ATO system sometimes fails to receive a signal to stop, forcing operators to manually respond by hitting a red button that brings the train to a halt. Even in a recent study involving an operational aircraft surveillance project, a decrement in performance was observed, though the effect was not as strong as those found in laboratory studies (Pigeau, Angus, O'Neill & Mack, 1995). In addition, submarine, helicopter, and surface ship operators highly ranked boredom or monotony as a stressor when they were asked to judge its impact on their overall performance as sonar operators (Wylie, Mackie, and Smith as cited in Mackie, 1987). All these findings, in providing support for the possible existence of a vigilance problem in the operational environment, highlight the significance of the role of human monitors even in this technological age.

Mackie (1987) is less harsh than Adams in criticizing vigilance research. He believes that we have done well in developing and testing various theoretical models in this area; however, a problem arises when one seeks the direct application of measures to remedy vigilance problems in operational or simulated operational tasks. In typical experimental or basic research, an arbitrary experimental task is selected and repeated manipulations of various parameters pertaining to the task has essentially resulted in a science of behaviour concerned with that particular task itself. Mackie (1987), cautioning against the repeated use of a particular task, believes that researchers now need to go a step further and to examine the effects of these same manipulations on more operationally realistic tasks.

Davies, Parasuraman and Craig, despite their contributions to the area of vigilance research, raise their own questions regarding the advances in the analysis of vigilance performance (Mackie, 1987). For example, Craig (1984) indicated that we have merely 'scratched the surface' of the underlying vigilance problem, suggesting that the monitoring task and the role of the human monitor has changed considerably since the days of Mackworth. Davies and Parasuraman (1982) also believe that the most significant advance in the area of vigilance research is yet to come, and is likely to be achieved through the investigation of situations that closely approximate the operational environment. As a remedy to this vigilance research situation, Weiner (1987) suggests that a greater effort to apply what we already know must take precedence over the quest for more research or for better-designed experiments. In providing direction to future researchers, Adams (1987) cited the use of semi-realistic tasks as the best way to proceed, adding that this is appropriate only if a system relevant problem can be identified.

Task Features and Complexity

The arguments presented above have been raised often. For example, Smith and Lucaccini (1969) not only questioned the applicability of vigilance research to practical problems, but also the parallelism of findings (especially vigilance decrement) across military monitoring and industrial inspection tasks. For example, Harris (1969) argued that findings on monitoring-type tasks (typically used in vigilance research) must be interpreted with caution when considering performance on other task types (for example, inspection, scanning, etc.). Monitoring and inspection tasks have a common vigilance element in that they require the detection of specified signals over an extended time period; yet, they differ in their task-specific features.

Task features have been used in vigilance research to describe the complexity of tasks (Davies & Parasuraman, 1982). For example, inspection tasks are described as being more complex than monitoring tasks because the responder or inspector is required to first interpret a specified quality standard, then compare a quality characteristic with the specified standard, make a decision about the conformity of the characteristic to the standard, and finally take action depending on the outcomes of the preceding responses in the sequence (Harris, 1969). On the other hand, relatively little judgement is required by monitors who observe a display for critical signals and make a response on the occurrence of such signals.

Vigilance tasks have often been classified as simple or complex, yet no explicit definition of complexity exists. In visual tasks, task complexity has often been manipulated through the alteration of certain display features or through the type of response required by the observer (Davies & Parasuraman, 1982). Whatever manipulation is chosen, task complexity has been studied extensively to reveal that it hinders vigilance performance, suggesting that complex tasks place a greater load on the performer than do simple tasks (for example, Harris, 1966; Adams, Humes, and Seivking, 1963). However, findings regarding complexity and different aspects of vigilance performance are inconsistent. For example, Koelega, Brinkman, Hendriks, and Verbaten (1989) found that overall performance (that is, percent of correct detections) on two equally complex and difficult tasks (based on participants' perceptions) was no different on either task, but a decline in performance was observed on only one of them. There are several interpretations of task complexity and its effects on performance decrement. When explained within a motivational frame-work, complexity is said to arrest vigilance decrement by replacing monotony with challenge. On the other hand, arousal theory maintains that difficult or complex tasks exert a strain on the attentional resources of the performer, resulting in lowered arousal, which progressively affects performance with time on task (Parasuraman, 1985).

No unitary theory has adequately explained much of the experimental evidence on performance decrement (Davies & Parasuraman, 1982). In discussing sensitivity decrement, Davies and Parasuraman (1982) conclude that so far the information-processing demands placed on the performer by different vigilance tasks appears to account for a large proportion of the available findings on decrement; however, they do not rule out other possible theoretical explanations such as effects of practice or memory load on sensitivity decrement, concluding that further work still needs to be done on these variables.

Task features, Information-processing, and Performance

In vigilance research, task complexity has generally been explained in terms of the processing demands imposed by the task feature(s) (Parasuraman, 1976). Some vigilance tasks are said to require more effort in the processing of stimuli than others. Effort, an aspect of attention that may be conceptualized as the capacity of the performer or the degree of conscious work invested in the task (Kahneman, 1973), is a frequently-used term in research on mental processing of dual tasks. Effort is generally used to describe the amount of resources demanded by a task where tasks requiring more processing (and consequently more effort) are said to exert a strain on the attentional resources of the performer such that performance efficiency decreases (Kahneman, 1973).

One task parameter, type of discrimination, has been identified as an important feature in the taxonomy of vigilance tasks and it is used to distinguish vigilance tasks in terms of their demand on attentional resources (or effort) (Davies & Parasuraman, 1982). Successive tasks generally involve the detection of a change in the intensity of light flashes or an increase in the deflection of a meter needle, where as simultaneous tasks involve the detection of a specified configuration in a complex pattern of vigilance tasks (Davies & Parasuraman, 1982). These tasks were originally differentiated by their underlying ability requirements (Levine, Romashko & Fleishman, 1973), but further examination of these tasks led to the identification of their information-processing requirements.

In successive tasks, responders must maintain a standard representation in working memory and compare successively presented stimulus configurations against that standard in order to make a discrimination. On the other hand, all information required to make a simultaneous discrimination is presented within the stimulus configuration itself. Therefore, successive tasks, because of their demand on memory, are said to exert a greater strain on the information processing resources of an individual than simultaneous tasks (Davies & Parasuraman, 1982). Performance over time on successive tasks, in comparison to that of simultaneous tasks, is said to suffer because of this greater expense of effort.

There has been general support for this information-processing distinction between vigilance tasks (Lanzetta, Dember, Warm & Berch, 1987; Parasuraman & Mouloua, 1987);

however, in independent studies on task-type (that is, successive versus simultaneous), the interacting influence of other task parameters on performance appears to be a consistent finding (Koelega et. al., 1989). In a recent meta-analysis, type of discrimination (that is, successive-simultaneous) was found to be insufficient in predicting sensitivity decrement (See, Howe, Warm & Dember, 1995). Decrement was found to be a function of other factors such as event-rate and stimulus-type as well. These findings have generally led researchers to propose a new taxonomy, adding another dichotomous (sensory-cognitive) category of vigilance tasks.

From an operational perspective such information is discouraging as operational tasks are so varied that the identification of task-specific features would probably lead to the placement of each task in a distinct category of its own. As Craig (1984) indicated, it is easy to gain an understanding of the vigilance process in laboratory settings because the task is simplified to the basic elements of concern; yet, in military and industrial settings, different types of stimuli occur within the constraints of a single task. Mackie (1987) also demonstrated how specific conditions imposed by a task and the task environment can have an impact on human responses, irrespective of whether they take place in the laboratory or in an operational setting. This leads us to question whether and how the nature of performance on tasks that differ on several dimensions is affected.

Identifying Features of the Tasks under Consideration

Inspection and monitoring tasks, as described by Harris (1969) and Smith and Lucaccini (1969), can be said to differ in their display features and response requirements. As the preceding discussion suggests, these two tasks (under consideration in this study) must be examined in detail to identify all features that can be anticipated to impact performance based on prior research. Therefore, this section examines the two vigilance tasks in terms of their task-specific features and response requirements. Both tasks have been used independently in previous studies as simulations of operational vigilance tasks (for example, Methot & Huitema, 1998; Mason & Redmon, 1992). A typical computer screen display for each of the following tasks is presented in Appendix A.

The monitoring task consists of two rectangular gauges placed vertically and each having the top and bottom sections marked as danger zones with a central neutral or noise zone. Signals, in the form of black arrows, are designated as critical when they appear besides any of the danger zones. Monitors were expected to make a response to critical signals by pressing on the space-bar of the attached computer key-board. Responses made in the absence of a critical signal were recorded as false alarms.

The inspection task was designed to simulate an assembly-line inspection of computer hard-disk drives. Each screen display presented a single disk-drive, containing components typical of any computer hard-drive. A missing component or critical signal was identified by comparing the given disk-drive to a sample representation devoid of defects. Inspectors were allowed to view the exemplar model to compare the given disk-drive by clicking on a 'view model' icon on the screen. Upon the identification of critical signals, the inspector was required to make an accept/reject decision by clicking on the attached computer mouse. Wrong reject decisions (that is, rejecting a disk-drive that had no defects) were recorded as false alarms. Important features of the monitoring and inspection task are discussed below.

Type of discrimination. The monitoring task requires simultaneous discrimination because signals are identified by their location in a clearly-marked danger zone. All information necessary to make a response is present within a given stimulus configuration. Individuals performing the monitoring task are required to press the space-bar on the keyboard whenever a signal is detected. In essence, responses (bar-pressing) can be made instantaneously, without substantial processing of information.

The quality control (QC) inspection task, on the other hand, requires successive discrimination because signals are identified after comparing the given stimulus configuration with a sample representation stored in memory. Responses on the inspection task require performers to accept or reject a disk-drive upon determining whether or not a defect was present within the given stimulus configuration by clicking on a mouse.

Stimulus Heterogeneity. This task feature is concerned with background or neutral events within which a signal is presented. For example, high similarity between background events and signals creates a homogenous stimulus event, whereas low similarity between the two creates a heterogeneous stimulus event. An observational comparison of the stimulus displays of the inspection and monitoring task (see Appendix A) indicates that the monitoring task affords greater stimulus heterogeneity than the QC inspection task. In other words, critical signals on the monitoring task are more noticeable and easily identifiable than those on the inspection task because the contrast between the signal and the stimulus configuration within which the signal appears is greater in the former task than in the latter.

Number of signals. In the monitoring task, individuals observed the stimulus configuration for a single signal type (an arrow in a marked danger zone), while on the QC inspection task, performers looked for 10 different signal types (different types of component defects).

Time Pressure. Both tasks impose a pressure on the individual to respond within a certain time frame. A delay in responding to a critical signal on the monitoring task results in a missed signal (if no response was made within 1.5 seconds after signal presentation) or a false alarm (if a response was made after 1.5 seconds following the onset of a signal), while a delay in responding on the QC inspection task resulted in a missed response.

Feature unique to Inspection task. The QC inspection task, designed to approximate an operational inspection task, allows performers to refer to the sample quality standard at any time during the task. This provision was based on the assumption that inspectors in the operational setting have access to a model stimulus during their work (Mason & Redmon, 1992).

Operational Relevance. Mackie (1977) used three dimensions in the classification of tasks in terms of their operational relevance: task-specific characteristics (for example, target to be detected and post-detection responses), the environment in which the task is performed, and temporal characteristics (for example, how often they are performed). Mackie's (1977) matrix of operational relevance reveals that the monitoring and inspection task also differ in their relevance to the operational setting.

Using this three-point scale classification, tasks with a rating of 1-1-1 are high on operational relevance, while those rated as 3-3-3 are low on operational relevance. The monitoring and inspection tasks under consideration in this study differ on the first dimension (i.e, task characteristics), receiving a rating of 3 (low) and 2 (moderate) respectively. In keeping with Adam's (1987) and Mackie's (1987) urge to use semi-realistic and operationally relevant tasks, the examination of other dimensions (environment and temporal characteristics) would have been ideal, but not without costs to the interpretation of findings. As seen in the operational relevance matrix, the degree of experimental control is lowered as operational relevance increases. Because this study is concerned with the task itself, the environmental and temporal dimensions were held constant and were not examined in this study.

Task features and Overall Performance

As seen in the examination above the two tasks differ on several parameters. According to Davies and Parasuraman (1982), the effects of the above independent task parameters (for example, stimulus heterogeneity, task-type) on vigilance performance are additive. The QC inspection task is more complex than the monitoring task, not only in the type of discrimination (successive) required, but in several other task characteristics as well. For example, greater stimulus homogeneity, the larger number of signals to be detected, and time pressure to respond add to the demands imposed by this successive task. All these features together create a task that demands greater effort than that required on the monitoring task, such that performance is likely to be poorer on the successive task.

Even though no single study has examined the combined effects of all the above factors on vigilance performance, findings from studies that examined combinations of at least two of these factors offer some support for the hypothesis that performance on tasks that demand greater effort is negatively affected. For example, from Lanzetta et. al's (1987) findings, the overall percent of accurate detections should be superior on the monitoring task (which is simultaneous and has stimulus heterogeneity) than on the inspection task (successive-homogeneous) because lesser effort is demanded on the former than on the latter.

Hypothesis 1: Overall performance (percent of correct signal detections) on the monitoring task will be greater than that on the QC inspection task.

Task features and Vigilance Decrement

As for performance decrement, which was not assessed in Lanzetta et. al's study, several hypotheses are plausible. As Parasuraman's (1985) multi-factor theory suggests, different theoretical models may be required in the examination of overall performance and performance decrement, although (as mentioned earlier) no single theory has been consistent in accounting for the latter. On one hand, performance can be expected to decline over time on the more complex task (that is, the inspection task) because processing requirements on this task exert a strain on the resources of the performer as time-on-task progresses. This proposition is supported by findings by Adams, Humes, and Sieveking (1963) where a within-session (three-hours) decrement was observed on a complex visual display task. However, in the absence of an experimental comparison with performance on a simple visual display task, it is difficult to confidently hypothesize that the opposite will hold true for the monitoring task.

On the other hand, sensitivity (a bias-free measure of performance decrement) was found to decline on both successive and simultaneous tasks when poorly discriminable signals were used, but when signals were highly or easily discriminable sensitivity declined only on the successive task and actually increased on the simultaneous task (Parasuraman and Mouloua, 1987). This suggests that sensitivity on the monitoring task (which presents easily discriminable signals in comparison to that on the inspection task) can be expected to remain the same or to increase with time on task, while sensitivity on the inspection task should decrease because it is successive and critical signals are not easily discriminable. However, it can also be argued that the discriminability of signals on the inspection task becomes easier as performers become familiar with the signals. In other words, as inspectors learn to recognize the types of defects (critical signals), they may find it easier to identify these signals and to respond to them accurately. Some support for this inherent learning of signals is provided by findings from the Mason and Redmon (1992) study (which used the same inspection task under consideration in this study) which show that participants viewed the sample stimulus frequently during the first 10 sessions, but hardly ever referred to it as the number of sessions progressed.

Loeb, Noonan, Ash, and Holding (1987) speculated how vigilance decrement may actually be avoided on complex tasks as participants learn efficient strategies for dealing with the task and how memory load may gradually be removed due to this learning process. However, they recognized that the validity of this proposition comes into question when after the learning of critical signals, the task becomes similar to any other simple vigilance task. In this view, sensitivity towards the latter portions of the inspection task can be hypothesized to be similar to that of the monitoring task. Sensitivity, which takes into account the proportion of target hits and false alarms, will be lower at the beginning of the inspection task (reflecting chance performance) relative to that towards the middle or end of the task. Participants may perform better towards the middle or end of the task than at the beginning because of the inherent learning that takes place, actually making the signals easier to recognize as time on task progresses.

In Mason and Redmon's (1992) discussion of performance on the inspection task, they mentioned how a particular aspect of the task (having four sample stimulus configurations presented on screen at the same time) may have aided participants' discrimination of signals by permitting a comparison across each configuration. This suggestion appears plausible; however, it could only have developed as time on task progressed. In other words, it could reflect one of the strategies developed by performers as they tried to efficiently meet some of the demands imposed by this task. As within-session decrement was not assessed in the Mason and Redmon study, objective evidence for either one of the above propositions remains lacking.

Hypothesis 2: Sensitivity will differ as a function of time on the QC inspection task, but not on the monitoring task. Estimates of sensitivity decrement on both tasks will be similar towards the latter portions of the task.

Hypothesis 3: The number of false alarms as a function of time will differ on both tasks. Specifically, false alarms will decrease with time on the QC inspection task.

Feedback and Performance

Feedback or knowledge of results (KOR), defined as knowledge of the quality or quantity of one's performance (Prue & Fairbank, 1981), has been demonstrated to have a positive effect on performance for monitoring and inspection tasks (for example, Mason & Redmon, 1992; Mackie, Wylie & Smith, 1994). However, there hasn't been a comparison of its effects across the two tasks. From an operational perspective, the study of feedback is valuable as it can be easily applied to a range of task-types. In addition, organizations frequently employ feedback because of its low cost, easy implementation, and positive effects (Prue &

Fairbank, 1981).

Peterson (1982) indicated that feedback (when it does work) has never been subjected to a detailed behavioural analysis of *why* it did work As a physical stimulus, feedback may function as a conditioned reinforcer, punisher or discriminative stimulus, and it is still not clear as to why it works in many cases. With respect to the present study, feedback may act as an antecedent to a response as well as a consequence for the accuracy of that response. Since the empirical analysis of feedback is not the focus of this study, no attempt was made to rationalize *why* feedback is hypothesized to positively influence overall performance on both tasks under consideration. Positive evidence for the immediate and quantitative nature of feedback on overall monitoring and inspection performance has been independently reported by Methot (under review) and Mason and Redmon (1992).

In addition to improvements in overall performance, Methot (under review) also demonstrated the positive effects of feedback on inter-individual consistency in performance at a monitoring task, where the provision of feedback resulted in smaller performance differences across individuals. Therefore, feedback was used in the present study to compare how its presence affects various aspects of vigilance performance across the monitoring and inspection task.

Hypothesis 4: The provision of feedback will serve to improve overall detections on both tasks.

Hypothesis 5: Inter-individual differences in performance across both tasks will be less when feedback is provided than when it is absent.

Individual Differences in Vigilance Performance

Individual differences in performance on vigilance tasks has been a consistent finding in past research (Davies & Parasuraman, 1982). There is accumulating support for the notion that these differences may be task-specific (for example, Parasuraman, 1976; Koelega et. al., 1989). Recently, Methot and Huitema (1998) found variability in performance on a monitoring task to be related to signal probability (another task dimension which has to do with the rate at which critical signals are presented), with higher probabilities generating consistent performance across individuals. Similarly, a correlation of performance across three different versions of the same task was found to be quite low, having only 12% of common variance (Stanislaw, 1995) between them.

While a number of parameters may influence the proportion of performance consistency, type of discrimination has been studied as one of the principal factors affecting performance across individuals. As previously discussed, type of discrimination, used in the taxonomic classification of vigilance tasks, distinguishes tasks in terms of the underlying ability required to perform the task (Levine et al., 1973). However, in the absence of a large-scale factor analytic study, Davies and Parasuraman (1982) remain speculative of whether this dichotomous dimension distinctly represents the necessary abilities required for efficient vigilance performance.

Predicting Vigilance Performance

Another dimension in the taxonomy called selective attention is described as the ability to perform a task in the presence of distracting stimulation without loss of efficiency (Levine, Romashko & Fleishman,1973). However, selective attention fails to distinctly differentiate between any two vigilance tasks (Davies & Parasuraman, 1982), suggesting that this ability may be mandatory for *all* vigilance tasks. Findings from a recent study (which employed the same monitoring task under consideration in this study) offer some support for this implication, where selective attention (a cognitive factor) was found to explain 16% of the variance in overall target hits (Darr, 1998). However, it failed to adequately account for false alarms and performance decrement.

Personality factors, especially extraversion, have been studied extensively in the prediction of vigilance performance. Introverts are said to outperform extraverts because of their ability to sustain attention over the length of the task (Eysenck, 1989). However, a recent meta-analysis on extraversion and vigilance reveals inconsistent findings (Koelega, 1992). Results from Darr (1998), which employed impulsivity (a component of extraversion) in the prediction of performance decrement and false alarms, suggests that impulsivity may be a better predictor of these aspects of vigilance performance than extraversion and/or selective attention. In addition, boredom susceptibility, which appeared promising in a previous unpublished study examining monitoring performance (Grant, 1997), was also examined along with impulsivity and extraversion. This procedure of using two different types of predictors (that is, personality and cognitive) in predicting various aspects of vigilance is consistent with Parasuraman's (1985) multi-factor approach in accounting for different aspects of performance on vigilance tasks.

Therefore, the present study attempted to replicate the findings of Darr (1998) to provide support for the usefulness of employing two kinds of predictors in predicting different aspects of vigilance performance. In the light of the discussions on vigilance performance and task specificity, this study also aimed at determining the applicability of these two predictors in predicting performance on the quality control inspection task.

Hypothesis 6: Overall performance (correct detections) on both tasks will be associated with measures of cognitive ability (selective attention).

Hypothesis 7: Performance decrement and false alarms on both tasks will be associated with impulsivity (a personality factor).

Summary of Hypotheses

Keeping in mind the previously-presented criticisms of vigilance research, the objective was not to design tasks according to parameters that afford optimal performance, but to examine tasks that bear more relevance to the operational setting, and to use what we already know in an attempt to predict and evaluate performance at such tasks. The purpose of this project, therefore, is to continue to contribute to vigilance research, taking into consideration some of the criticisms and issues raised by experts in this area. It is an attempt to apply what we already know to determine the applicability of such information to tasks that differ in their relevance to the operational setting.

This study examined performance across two different vigilance tasks in consideration of their task-specific features and evaluated the effects of an operationally-relevant independent variable on performance at these two tasks. It also attempted to replicate findings from a pilot study concerned with the prediction of performance on a monitoring task using two different predictors, and also aimed at determining the applicability of these predictors to performance on inspection tasks.

Given that vigilance performance is affected by the probability at which critical signals are presented to the observer (Methot & Huitema, 1998), signal probability was held constant across the monitoring and inspection tasks. Specifically, the probability of occurrence for each critical signal was set at 0.05 on both tasks in an attempt to approximate operational settings where the likelihood of critical signals is generally low. From Methot and Huitema's (1998) comparison of monitoring performance across three levels of signal probability (0.01, 0.04 and 0.12), 0.05 appears to be an intermediate and acceptable level such that the task is not too easy (thus avoiding possible ceiling effects) nor too difficult. With respect to past manipulations of probability on the inspection task, Mason and Redmon (1992) employed an extremely high signal probability level (0.50) in their examination of performance on the quality control (QC) inspection task.

Hypothesis 1. Overall performance (percent of correct signal detections) on the monitoring task will be greater than that on the QC inspection task.

Hypothesis 2. Sensitivity (using a non-parametric computation) will differ as a function of time on the QC inspection task, but not on the monitoring task. Estimates of sensitivity decrement on both tasks will be similar towards the latter portions of the task.

Hypothesis 3. The number of false alarms as a function of time will differ on both tasks. Specifically, false alarms will decrease with time on the QC inspection task.

Hypothesis 4. The provision of feedback will serve to improve overall detections on both tasks.

Hypothesis 5. Inter-individual differences in performance across both tasks will be less when feedback is provided than when it is absent.

Hypothesis 6. Overall performance (correct detections) on both tasks will be associated with measures of cognitive ability (selective attention).

Hypothesis 7. Performance decrement and false alarms on both tasks will be associated with impulsivity (a personality factor).

Method

Participants

Sixty-eight students, recruited from the general university population, participated in this study for a small fee. Those who were entitled to receive bonus points towards a psychology course were awarded two points in addition to the financial compensation. All participants had normal or corrected-to-normal vision.

Materials/Instrumentation

Computer simulations of the monitoring and inspection vigilance tasks and all paper-and-pencil tests, administered in this study, are individually described below.

Monitoring Task. This task, modified to suit the goals of this study, was similar to that used in Methot and Huitema (1998). Written in HyperCard (version 2.0.2) for Macintosh, a typical screen display for this task (see Appendix A) consisted of two vertical rectangular gauges, each divided into three sections (two danger and one noise). Participants were required to press the space-bar on the keyboard whenever an arrow appears beside any of the danger zones. The program was designed to present a short instructional tutorial prior to the task and to record responses during the entire vigil. Correct responses were recorded as hits, while incorrect responses (key presses in the absence of a signal) were recorded as false alarms.

Inspection Task. This task was similar to that employed by Mason and Redmon (1992), but was modified to suit the purposes of this study. Written in HyperCard (version 1.2.2) software, the task was designed to present schematic visual representations of a computer hard disk-drive (see Appendix A). Each representation consists of eight components (a voltage regulator, a pair of screws, a pair of soldered memory chips, a coprocessor chip, a fuse, a processor chip, six resistors, and a central screw of storage disks). Each screen presented one disk-drive representation at a time, each having an "accept" and a "reject" icon below it. Another icon labelled "view model", which allowed participants to view a sample disk-drive representation (with no errors), was located towards the top, extreme right of the screen. Participants were required to make a decision about whether each disk-drive representation should be accepted or rejected. Rejecting a defective disk drive (that is, one with a missing component) contributed to hit performance, where as rejecting a non-defective disk drive contributed to a performer's false alarm-rate. Disk-drives containing an error were required to be rejected, while those with no error were retained or accepted. The program was designed to present a short tutorial session (to familiarize participants with the defective disk components or signals to be detected) prior to the task and to record quality control responses during the entire task.

Digit Symbol. The revised Wechsler Adult Intelligence Scale (WAIS-R) has been widely researched as a valid and reliable measure of general intelligence. The Digit Symbol, a subtest on the WAIS-R, was used as a measure of selective attention in this study. The Digit Symbol task is a 90-second paper-and-pencil task requiring subjects to match as many number and symbol combinations within the specified time limit (Wechsler, 1981). It assesses the capacity for sustained attention, effort and concentration, and low scorers tend to have poorer mental alertness than high scorers (Groth-Marnat, 1990).

Impulsivity measure. The Emotionality, Activity, Sociability and Impulsivity (EASI -III, Buss & Plomin, 1975) scale was used to measure impulsivity. This measure was chosen because it appears to contain a well-researched representation of the impulsivity construct by its developers. Buss and Plomin (1975) originally developed the EASI-I, which measured impulsivity primarily in terms of inhibitory control. Further research led to the conceptualization of impulsivity as a multi-dimensional construct, leading to the development of the EASI-III which contains four facets: inhibitory control, decision time, sensation seeking and persistence. The EASI scales have been employed in many studies (for example, Windle, 1989; Gerbing, Ahadi & Patton, 1987) aimed at exploring inter-inventory comparisons, and it appears to correlate appropriately with other related measures such as the Revised Dimensions of Temperament Survey (DOTS-R), Eysenck's Personality Inventory (EPI), and the I 5 and I 7 scales.

The EASI-III is a paper-and-pencil questionnaire that requires individuals to respond to 50 statements about their general temperament (Buss & Plomin, 1975). The rating scale on the EASI-III was modified (from its existing format) by replacing the two-anchor rating scale (1 = a little, and 5 = a lot) with one that has five anchors (1= strongly disagree, 2= disagree, 3= neutral, 4= agree, and 5= strongly agree). This was done mainly to improve clarity and applicability to all items.

Sensation Seeking Scale (SSS), Form V. Zuckerman, Kolin, Price and Zoob (1964) originally developed the SSS to measure individual differences in optimal levels of arousal and stimulation. The scale contains four facets: thrill and adventure seeking, experience seeking, disinhibition, and boredom susceptibility. Only items pertaining to the boredom susceptibility facet were administered in the present study. As this facet provides a representation of a person's aversion to repetition, routine and restlessness when phenomena are unchanging (Zuckerman, Eysenck & Eysenck, 1978), it appeared to bear most relevance to the present study. Form V was developed from form IV, but each facet contained only 10 items as opposed to the previous 15 or 18. Cross-cultural (English and American samples) and gender studies conducted by the developers of the scale reveal that form V does slightly better than form IV on cross-cultural stability. Even though there were differences between the sexes, these differences were limited to the thrill-seeking and disinhibition sub-scales only (Zuckerman, Eysenck & Eysenck, 1978). Internal consistency reliability coefficients for the boredom susceptibility sub-scale ranged between 0.56 and 0.65 for the different groups compared.

Revised NEO Personality Inventory. The short form of the NEO PI-R (Form S), which measures five facets of personality, was used in this study. The facets of primary interest were Extraversion and Conscientiousness. The NEO PI-R has been well-researched, and internal consistency reliability estimates for individual facets range between 0.56 and 0.81. Test retest reliability estimates (reported in Costa, Jr. & McCrae, 1991) also range between 0.66 and 0.92. The NEO PI-R facets have also been demonstrated to have convergent validity with other measures of personality such as the Eysenck Personality Inventory (EPI) (see Eysenck & Eysenck, 1964).

Subjective measure. A subjective measure was designed to assess task complexity as perceived by the participants. The self-report three-item questionnaire also asked participants about the amount of effort expended on each task (see Appendix B).

Procedure

All participants individually completed both vigilance tasks (monitoring and inspection), presented in a counterbalanced order, over two experimental sessions. The sessions were scheduled on two different days, and were conducted in a quiet room which contained a desk, a chair, a Macintosh computer, mouse and keyboard. Prior to introducing the first vigilance task (that is, in the first session) participants completed the NEO PI-R, followed by the Digit Symbol sub-test and the Boredom Susceptibility sub-scale of the SSS . Upon completion of these questionnaires, participants were exposed to either the monitoring or the inspection tutorial (pre-determined using random assignment). Following the tutorial, any queries pertaining to the task were answered, and participants were then directed to begin the actual task. Those participants that completed the first session were scheduled to come in on a second day. The second session began with the administration of the EASI-III, followed by the second vigilance task, which was introduced in the same manner as previously described. Upon completion of the second vigilance task, participants were presented the subjective measure, and were then compensated for completing the entire experiment.

In an attempt to assess the relevance of present tasks to the operational setting, a meeting with the quality control director at a local electrical company was arranged The quality inspector concluded that the inspection task compares relatively well to that performed by assembly line circuit board inspectors at that company. In addition to scanning for missing components,

components, inspectors are also required to inspect for size and orientation of elements on a circuit board. A typical part examined by an inspector in the electronics setting contains approximately 120 elements (personal communication, July 1999).

Results

Pre-analysis screening

A preliminary screening of the data led to the deletion of two cases: one, due to failure to complete the entire experiment, and two, unusual performance profile on the monitoring task (participant fell asleep and did not respond during entire vigilance task). Therefore, a total of 66 cases were retained in the final analyses. False alarm rates across each 10-minute time sessions for one participant, who failed to follow instructions, were unusually high and were replaced with the average rate obtained for each 10-minute interval. Technical difficulties, encountered on the monitoring task during the initial phases of the experiment, resulted in the loss of false alarm estimates for 14 participants. However, these missing values were replaced using one of the procedures (regression) outlined in Tabachnick and Fidell (1996), described to be more objective than using the grand mean. False alarm estimates were obtained using a regression equation with proportion of target hits as a predictor. As both these variables are significantly correlated ($r = -0.57$, $p < 0.001$), this approach was deemed appropriate, and it yielded an estimate of the false alarm-rate for each participant based on their proportion of total hits on the monitoring task.

Dependent Measures

Target hits. On the monitoring task, overall target hits were the total percentage of arrows or critical signals detected (that is, pressing the space bar whenever a signal entered the danger zone). Target hits on the inspection task constituted the number of defective disk-drive screens that were rejected, divided by the total number of defective screens presented during the task.

False alarms. The rate of false alarms on the monitoring task was estimated as the proportion of responses (bar presses) made in the absence of a critical signal to the total number of non-critical stimulus changes. False alarm rates on the inspection task were calculated as the number of non-defective disk-drive screens that were wrongly rejected, divided by the total number of non-defective screens that were correctly retained or accepted.

Decrement. Performance decrement for each participant was calculated using the proportion of hits recorded at each 10-minute time interval on the monitoring task. Therefore, hit performance was recorded at six time intervals across the length of the hour-long monitoring task. As the length of the inspection task could not be held constant for all participants (that is, although the task was designed to stop after an hour, most participants were able to scan all 600 screens before the hour was up), the total number of scanned screens for each participant were divided by six, and the number of hits within each of the six sets were used in the estimation of performance decrement. Performance decrement across the length of each of the vigilance tasks was then calculated using the following linear model, earlier identified by Methot and Huitema (1998):

$$Y_{i,t} = \beta_o + \delta_i + \varepsilon_{i,t}$$

where,

$Y_{i,t}$ is the hit score for participant i at dimension t,

β_o is the regression intercept,

δ_i is the decrement coefficient that describes the average number of units that response measure Y changes per unit of time, and

$\varepsilon_{i,t}$ is the error on the participant i at dimension t.

Within session Variability. According to Methot and Huitema (1998), inconsistent performance, for each participant, across the length of the vigilance task was found to be best represented by the residual variance obtained using the linear regression model above.

Perceptual Sensitivity. This measure was obtained using the following non-parametric computation identified by Grier (1971):

$$A' = \tfrac{1}{2} + [(H - FA)^* (1 + H - FA)] / [(4H) * (1 - FA)]$$

where, H = proportion of hits and FA = false alarm rate. The resulting sensitivity index (A′) increases from 0.50, reflecting chance performance to 1.0, reflecting perfect performance.

Statistical Analyses

Regression. The usefulness of each predictor was assessed by regressing each dependent measure (target hits, decrement, false alarms, and variability in performance over time) onto all predictors simultaneously. However, in an attempt to control for the experimental manipulation of this study, a hierarchical regression analysis was necessary. In other words, group membership was entered on step I followed by all the predictor variables (boredom susceptibility, digit symbol, impulsivity and extraversion) simultaneously on step II. With respect to the inspection task, it was also necessary to statistically control for the length of the inspection task. Therefore, in this case predictor variables were entered on step III. The average length of the inspection task was 44.47 minutes ($\underline{SD}$ = 9.39).

Repeated measures ANOVA. This approach was used to analyse differences in the dependent measures across task-type (the within-subjects factor) and across feedback conditions (the between-subjects factor). Because several dependent measures were recorded on each task, this analysis is sometimes referred to as the doubly multivariate repeated measures model.

Chi-square. Data from the self-report (subjective) measure of task complexity and effort, being categorical in nature, were analysed using a chi-square analysis.

Predictors of Vigilance Performance

The regression coefficients of all predictor variables on each performance measure for both tasks are presented in Tables 1a and 1b. Performance measures on the monitoring task are significantly inter-correlated (Tables 2a and 2b); however, each measure is discussed separately in an attempt to facilitate interpretation of the direction and magnitude of the obtained associations.

Overall target hits. Of the four predictors used to predict this measure of performance, impulsivity ($\beta = -0.35$, $p < 0.01$) and extraversion ($\beta = 0.28$, $p < 0.05$) emerged as the only significant predictors of overall proportion of hits on the monitoring task. However, none of the four predictors were useful in predicting hits on the inspection task. Even though impulsivity is a facet of extraversion, the two variables are weakly correlated ($r = 0.145$, $p > 0.05$).

Decrement. None of the predictors were significantly associated with this measure of performance on either of the tasks, although the association of impulsivity and decrement on the monitoring task appears to be approaching significance ($\beta = -0.23$, $p = 0.07$). This relationship is also in the expected direction; low impulsivity scores are related to stable or increasing hit performance over time on the monitoring task.

False alarms. False alarm rate on the monitoring task was best predicted by impulsivity again ($\beta = 0.34$, $p < 0.01$), while boredom susceptibility ($\beta = 0.25$, $p < 0.05$) and the digit symbol task (β= -0.27, p <0.05) were significantly associated with false alarms on the inspection task. As expected, high impulsivity is associated with higher false alarm rates on the monitoring task. With respect to this performance measure on the inspection task, those scoring high on the boredom susceptibility measure (indicative of being easily susceptible to boredom) had higher false alarm rates than those scoring low on this measure. Similarly, the negative association between digit symbol scores and false alarm rates suggests that lower scores on this measure of selective attention are related to higher false alarm rate on the inspection task. Both these obtained associations are consistent with a priori expectations, providing some insight into this measure of performance.

Within-session variability. Impulsivity (β = 0.28, p <0.05) and extraversion (β = -0.34, p <0.01) predicted within-session variability on the monitoring task, but not on the inspection task. Again, the association between impulsivity and performance variability across time on the monitoring task is consistent with that expected; those scoring low on impulsivity were more likely to perform consistently (lower variability) throughout the task. Even though the obtained effect size for extraversion is larger, the negative association is inconsistent with past research findings.

Perceptual Sensitivity. Perceptual sensitivity on the monitoring task was best explained by impulsivity (β= -0.40, p < 0.01). No significant associations were obtained for this performance measure on the inspection task. Because perceptual sensitivity is a function of target hits and false alarm rate, its obtained association with impulsivity is not novel. The direction of this association suggests that low impulsivity is associated with higher perceptual sensitivity to the monitoring task.

Table 1a

Regression of predictor variables onto performance measures on the Monitoring Task

Monitoring Task -Target Hits			
Predictor variables	β	T	p
Impulsivity (EASI)	-0.345	-2.809	0.006**
Extraversion (NEO)	0.278	2.367	0.021*
Boredom Susceptibility (SSS)	-0.073	-0.614	0.541
Digit Symbol (WAIS-R)	-0.047	-0.401	0.689
Monitoring Task – Decrement			
Impulsivity (EASI)	-0.234	-1.831	0.072
Extraversion (NEO)	0.149	1.220	0.227
Boredom Susceptibility (SSS)	0.078	0.639	0.525
Digit Symbol (WAIS-R)	-0.004	-0.035	0.971
Monitoring Task - False Alarms			
Impulsivity (EASI)	0.340	2.792	0.007**
Boredom Susceptibility (SSS)	0.181	1.521	0.133
Extraversion (NEO)	-0.197	-1.697	0.095
Digit Symbol (WAIS-R)	-0.114	-0.981	0.330
Monitoring Task - Within-session Variability			
Extraversion (NEO)	-0.335	-2.892	0.005**
Impulsivity (EASI)	0.280	2.310	0.024*
Digit Symbol (WAIS-R)	-0.100	-0.864	0.391
Boredom Susceptibility (SSS)	0.099	0.840	0.404

Monitoring Task - Perceptual Sensitivity			
Impulsivity (EASI)	-0.395	-2.830	0.006**
Extraversion (NEO)	0.182	1.353	0.182
Boredom Susceptibility (SSS)	-0.151	-1.111	0.272
Digit Symbol (WAIS-R)	-0.011	-0.087	0.930

** $p < 0.01$

* $p < 0.05$

Table 1b

Regression of predictor variables onto performance measures on the Inspection Task

Predictor variables	β	T	p
Inspection Task -Target Hits			
Boredom Susceptibility (SSS)	-0.099	-0.805	0.424
Digit Symbol (WAIS-R)	0.097	0.819	0.416
Impulsivity (EASI)	0.059	0.474	0.637
Extraversion (NEO)	-0.013	-0.110	0.912
Inspection Task - Decrement			
Extraversion (NEO)	0.134	1.085	0.282
Impulsivity (EASI)	-0.092	-0.710	0.480
Boredom Susceptibility (SSS)	0.085	0.694	0.490
Digit Symbol (WAIS-R)	-0.081	-0.637	0.526
Inspection Task - False Alarms			
Digit Symbol (WAIS-R)	-0.268	-2.341	0.022*

Boredom Susceptibility (SSS)	0.250	2.093	0.040*
Extraversion (NEO)	-0.086	-0.746	0.458
Impulsivity (EASI)	0.073	0.610	0.544
Inspection Task - Within-session Variability			
Digit Symbol (WAIS-R)	0.132	1.036	0.304
Extraversion (NEO)	-0.086	-0.668	0.506
Boredom Susceptibility (SSS)	-0.074	-0.558	0.578
Impulsivity (EASI)	-0.026	-0.199	0.842
Inspection Task - Perceptual Sensitivity			
Impulsivity (EASI)	0.080	0.594	0.553
Digit Symbol (WAIS-R)	0.074	0.574	0.568
Extraversion (NEO)	-0.070	-0.523	0.603
Boredom Susceptibility (SSS)	-0.045	-0.336	0.738

* $p < 0.05$

Table 2a

Association between Predictor and Monitoring Performance variables (after controlling for group, N=51)

	H	FA	D	V	S	I	E	BS	DS
H	1.00	-0.52 ***	0.68***	-0.75***	0.66***	-0.34*	0.16	-0.13	-0.20
FA	-	1.00	-0.41**	0.59***	-0.95***	0.36*	-0.11	0.28*	-0.04
D	-	-	1.00	-0.46**	0.50***	-0.26	0.08	-0.21	0.03
V	-	-	-	1.00	-0.61***	0.21	-0.08	0.22	0.12
S	-	-	-	-	1.00	-0.39**	0.10	-0.21	-0.03
I	-	-	-	-	-	1.00	0.18	0.19	0.09
E	-	-	-	-	-	-	1. 00	0.06	-0.02
BS	-	-	-	-	-	-	-	1.00	-0.12
DS	-	-	-	-	-	-	-	-	1.00

* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

H = Target hits
FA = False Alarms
D = Decrement
V = Variability
S= Sensitivity

I = Impulsivity
E = Extraversion
BS = Boredom Susceptibility
DS = Digit Symbol (Selective Attention)

Table 2b

Association between Predictor and Inspection Performance variables (after controlling for group and length of task, N= 66)

	H	FA	D	V	S	I	E	BS	DS
H	1.00	-0.06	0.12	0.49***	0.92***	0.05	0.00	-0.10	0.11
FA	-	1.00	-0.06	-0.22	-0.19	0.10	-0.09	0.29**	-0.29 **
D	-	-	1.00	0.10	0.14	-0.09	0.13	-0.11	0.09
V	-	-	-	1.00	0.66***	-0.04	-0.08	-0.09	0.13
S	-	-	-	-	1.00	0.06	-0.01	-0.04	0.09
I	-	-	-	-	-	1.00	0.13	0.27 *	0.11
E	-	-	-	-	-	-	1.00	0.03	0.06
BS	-	-	-	-	-	-	-	1.00	-0.06
DS	-	-	-	-	-	-	-	-	1.00

* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

H = Target hits
FA = False Alarms
D = Decrement
V = Variability
S= Sensitivity

I = Impulsivity
E = Extraversion
BS = Boredom Susceptibility
DS = Digit Symbol (Selective Attention)

Task-type differences

There was a significant multivariate effect on the within-subjects factor, task-type (Wilks' $\lambda = 0.048$, $p < 0.001$). The accompanying univariate tests (Table 3a) show that monitoring and inspection task performance differ significantly on all the recorded measures, except decrement.

Table 3a

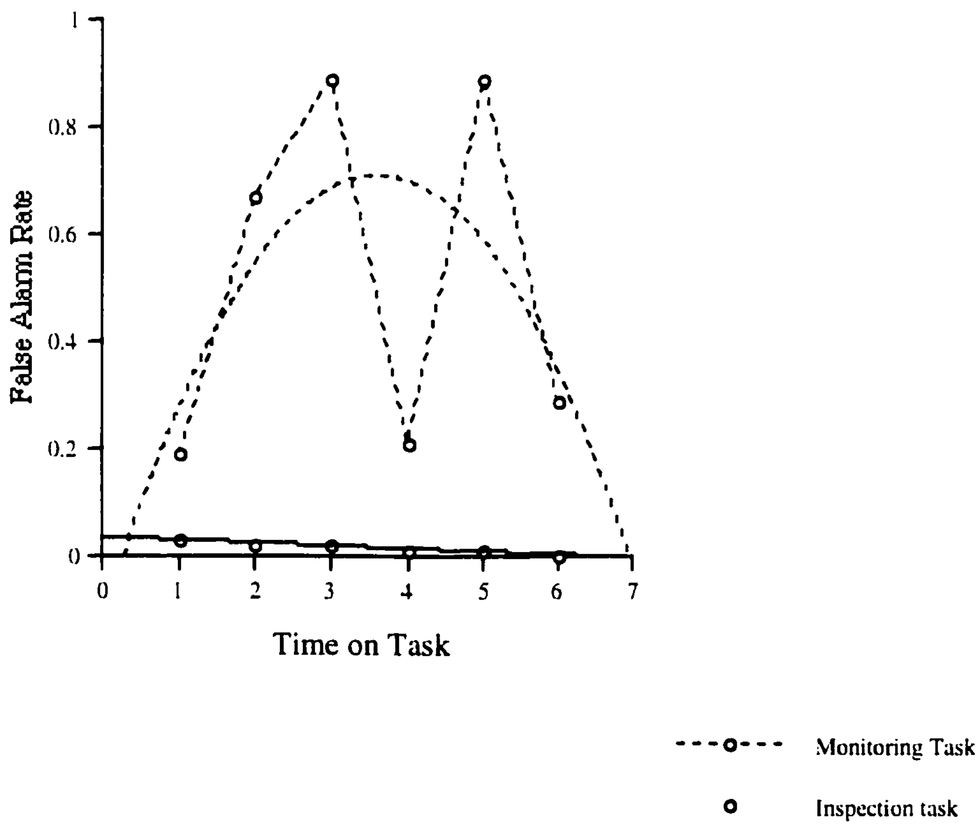
Summary of repeated-measures analyses showing the effects (univariate) of task-type and feedback on vigilance performance measures

Source	df	F	η^2	Obs. Power	p
Within-subjects (task-type)					
Target Hits	1	810.796	0.927	1.00	0.000***
False Alarms	1	33.828	0.346	1.00	0.000***
Decrement	1	0.765	0.120	0.14	0.385
Variability	1	37.826	0.371	1.00	0.000***
Between-subjects (feedback)					
Target Hits	1	5.233	0.076	0.615	0.025*
False Alarms	1	0.040	0.001	0.054	0.842
Decrement	1	0.222	0.003	0.075	0.639
Variability	1	0.444	0.007	0.101	0.508
Task-type X Feedback					
Target Hits	1	5.922	0.085	0.669	0.018*
False Alarms	1	0.064	0.001	0.057	0.801
Decrement	1	1.296	0.020	0.202	0.259
Variability	1	2.306	0.035	0.322	0.134
Error	64				

* $p < 0.05$ *** $p < 0.001$

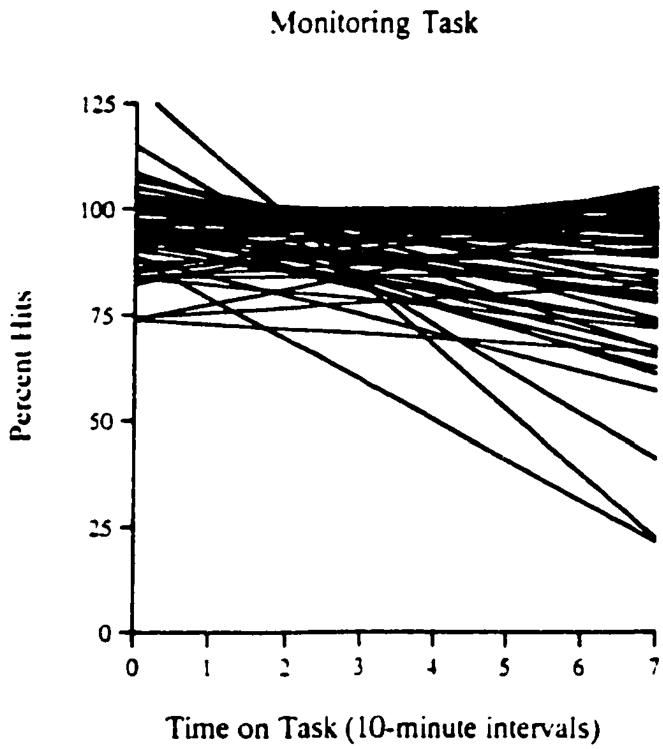
Descriptive information (Table 4) suggests that participants performed better on the monitoring task than on the inspection task in terms of overall performance, where percentage of hits on the monitoring task ($M = 90.48$, $SD = 8.88$) was significantly higher ($F(1,64) = 810.80$, $p < 0.001$) than on the inspection task ($M = 32.51$, $SD = 18.13$). However, the false alarm-rate on the monitoring task ($M = 0.25$, $SD = 0.33$) was higher than ($F(1, 64) = 33.82$, $p < 0.001$) on the inspection task ($M = 0.01$, $SD = 0.03$). A plot of the average false alarm-rate across time suggests that false alarm-rates on the inspection task were generally low and stable over time on task, but varied greatly over time on the monitoring task. The false-alarm rate over time on the monitoring task may be best described as being non-linear (see Figure 1). The obtained relationship suggests that the number of false alarms made on the monitoring task were lower in the beginning and at the end of the task, but tended to be higher during the middle of the task. However, in the absence of a test for linearity, this suggestion remains tentative. It is possible that the within session error variance is so high at this low signal probability that any linearity is masked. A linear fit has shown to be best at higher signal probability levels (e.g., Methot & Huitema, 1998).

Figure 1: False Alarm rate over time on the Monitoring and the Inspection Task



Consistent with the larger standard deviation obtained for hit performance on the inspection task, within-task performance varied substantially on this task. As seen in Figure 2, participants tended to perform, on average, similarly to one another on the monitoring task, but differed greatly from each other in their performance on the inspection task, $F(1, 64) = 37.83$, $p < 0.001$. Even though the average decline in performance over time on each task did not differ significantly from each other, average decrement on the monitoring task ($M = -1.54$, $SD = 3.03$) was higher than on the inspection task ($M = -0.83$, $SD = 6.50$). When hit performance within each of the six time intervals was averaged across participants, and regressed across time using the previously described linear model, there appears to be a considerable difference in the standardized slope coefficients for each task-type ($\beta = -0.85$, $p < 0.05$ for the monitoring task, and $\beta = -0.34$, $p > 0.05$ for the inspection task), suggesting that the decline in performance as time-on-task progressed was greater on the monitoring task than on the inspection task (see Figure 3).

Figure 2: Performance Variability within each Vigilance Task



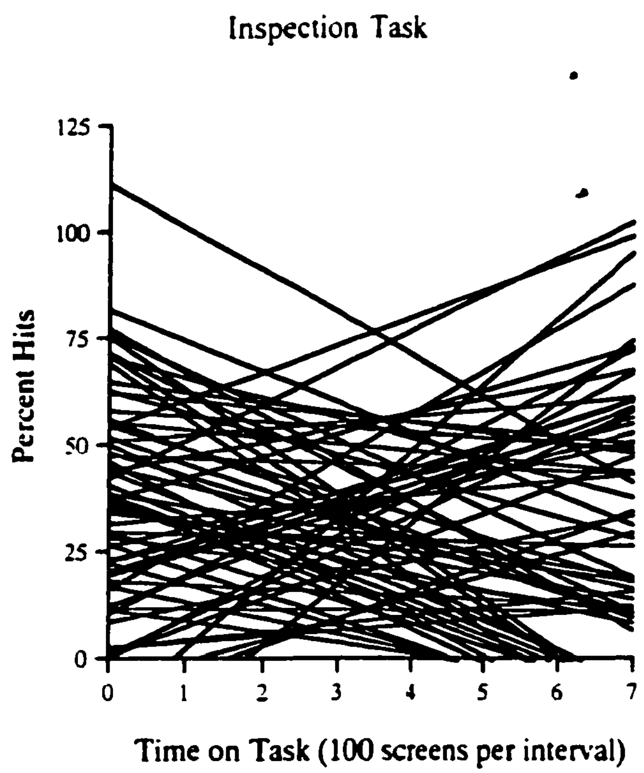
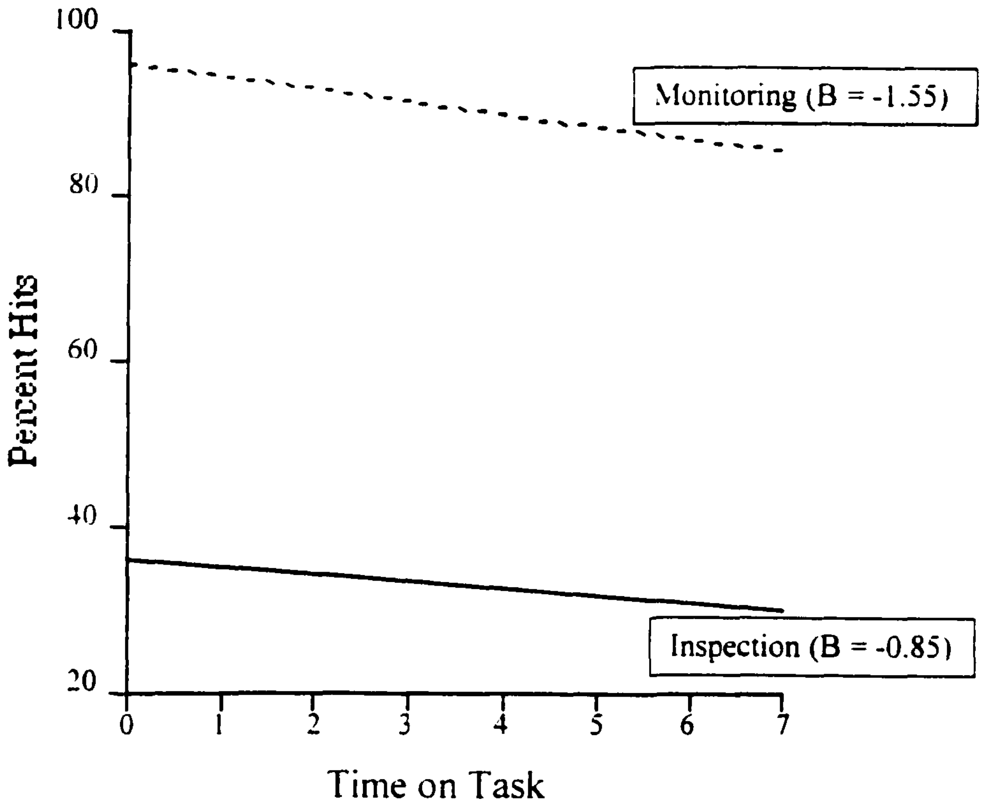


Table 4

Average performance measures for each vigilance task and feedback condition

	Monitoring		Inspection	
Overall (N = 66)	M	SD	M	SD
Target Hits	90.478	8.878	35.514	18.139
False Alarms	0.252	0.333	0.014	0.036
Decrement	-1.537	3.025	-0.831	6.497
Variability	119.311	265.876	596.580	535.759
Sensitivity (N = 51)	0.727	0.283	0.582	0.073
Feedback Condition (N = 33)				
Target Hits	89.840	9.726	26.922	19.369
False Alarms	0.253	0.371	0.005	0.007
Decrement	-0.852	2.542	-1.065	6.225
Variability	155.302	353.105	514.736	393.054
Sensitivity (N = 25)	0.699	0.318	0.572	0.075
No Feedback Condition (N = 33)				
Target Hits	91.117	8.041	38.106	15.122
False Alarms	0.251	0.295	0.024	0.049
Decrement	-2.221	3.341	-0.596	6.846
Variability	83.319	127.415	678.424	644.002
Sensitivity (N = 26)	0.755	0.247	0.591	0.072

Figure 3: Performance Decrement across time on each Task



Perceptual sensitivity, which was not included in the previous analysis (because of a large number of missing values) was examined separately (Table 3b), but with a slightly smaller sample size ($N = 51$). Average sensitivity differed significantly on both tasks, $F(1, 49) = 13.362$, $p < 0.01$, with it being higher on the monitoring task ($M = 0.73$, $SD = 0.28$) than on the inspection task ($M = 0.58$, $SD = 0.07$). With respect to sensitivity shifts over time on task, a regression analysis of the averaged sensitivity estimates (across each time interval and across participants) onto the six time intervals shows that sensitivity declined more rapidly on the monitoring task ($\beta = -0.458$) than on the inspection task ($\beta = -0.095$). These estimates suggest that sensitivity to the inspection task, though low, remained almost stable during the length of that task, but it declined over time on the monitoring task.

Table 3b

Summary of repeated-measures showing the effects of task-type and feedback on perceptual sensitivity

Source	df	F	η^2	Obs. Power	p
Within-subjects (task-type)					
Sensitivity	1	13.362	0.214	0.948	0.001**
Between-subjects (feedback)					
Sensitivity	1	0.779	0.016	0.139	0.382
Task-type X Feedback					
Sensitivity	1	0.218	0.004	0.074	0.642
Error	49				

** $p < 0.01$

<u>Differences across feedback conditions</u>

The significant task-by-group effect ($F(1, 64) = 5.92$, $p < 0.05$), which was evident only for target hits, suggests that feedback affected hit performance differently across the two tasks. However, the percentage of hits detected on both tasks was lower when feedback was given (Figure 4) than when feedback was absent. This effect is more pronounced on the inspection task (feedback : M = 26.92, SD = 19.37; no feedback: M = 38.11, SD = 15.12) than on the monitoring task (feedback: M = 89.84, SD = 9.72; no feedback: M = 91.12, SD = 8.04). Even though no other significant effects were found across feedback conditions, a visual examination of the average decrement plot (see figure 5) suggests that when feedback is present, average decrement on the monitoring task is lower (M = -0.85, SD = 2.54) than when feedback is absent (M = -2.22, SD = 3.34), but the effect is reversed on the inspection task where average decrement is lower when feedback is absent (M = -0.60, SD = 6.84) compared to when feedback is present (M = -1.06, SD = 6.22).

Figure 4: Average Hit Performance for each Vigilance Task and Feedback Condition

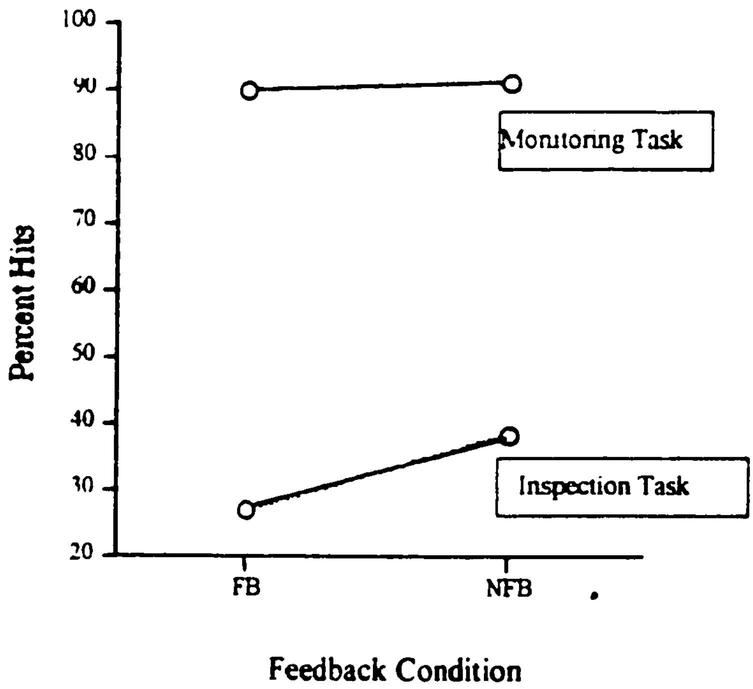
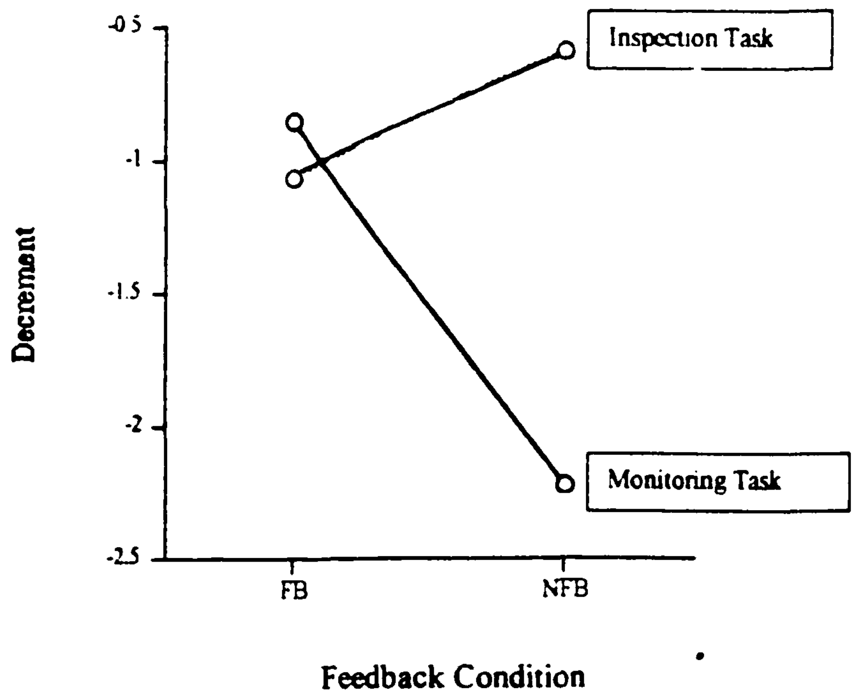


Figure 5: Average Decrement for each Vigilance Task and Feedback Condition



Subjective Measure

A significantly greater number of participants (80.3%) stated that the inspection task was more complex than the monitoring task ($\chi^2 = 107.69$, $p < 0.001$, Table 5). Approximately 17% said that the monitoring task required less effort as time passed on, 42.4% said that the task required more effort as time passed on, and 40.9% indicated that the task required no more effort at any particular time on the task. Therefore, an equal majority of participants ($\chi^2 = 0.018$, $p > 0.89$) agreed that the monitoring task either required more effort as time passed on or required the same amount of effort throughout. With respect to the inspection task, a significantly greater number of participants (68.2%) felt that the inspection task required less effort as time passed on ($\chi^2 = 13.78$, $p < 0.001$), while 24.2% felt that it required more effort with the passage of time, and only 7.6% felt that the task did not require any more effort at any particular time on the task. This distribution of responses was no different across the two feedback conditions ($p > 0.05$).

Table 5

Distribution on responses on the Subjective Measure

Item	Observed Percent
Complexity	
Monitoring task more complex	6.1
Inspection task more complex	80.3
Both tasks equally complex	7.6
None of the tasks were complex	6.1
Monitoring Task	
Required less effort as time passed on	16.7

Required more effort as time passed on	42.4
Required no more effort at any particular time	40.9
Inspection Task	
Required less effort as time passed on	68.2
Required more effort as time passed on	24.2
Required no more effort at any particular time	7.6

Post Hoc Analyses

The decline in the number of times the sample stimulus was viewed on the inspection task provides some evidence for a learning effect. In other words, the first half of the inspection task may have served as a training phase for participants, giving rise to the possibility that performance outcomes on the second half of the inspection task may have been confounded by this training phase. As a result, task-type and feedback effects were reanalysed using performance data from the latter portion of the inspection task (that is, the last three sessions). Results suggest that task-type differences in performance (hits, false-alarm rate, and sensitivity) continue to exist even after attempting to control for training phase on the inspection task (Table 6). Task by group effect for hit performance is no longer significant, but may be a result of reduced power due to the inclusion of sensitivity (which had 14 missing cases). Examination of average hit performance on the inspection task reveals that hits continue to remain low when feedback is given (M = 28.33, SD = 25.30) than when no feedback is given (M = 39.50, SD = 21.06).

Table 6

Post hoc repeated-measures showing the effects of task-type and feedback on hit performance, false alarm rate , and sensitivity

Source	df	F	η^2	Obs. Power	p
Within-subjects (task-type)					
Hits	1	340.77	0.874	1.00	0.000***
False Alarms	1	25.41	0.342	0.99	0.000***
Sensitivity	1	11.69	0.193	0.918	0.001**
Between-subjects (feedback)					
Hits	1	2.956	0.057	0.392	0.092
False Alarms	1	0.011	0.000	0.051	0.916
Sensitivity	1	0.735	0.015	0.134	0.134
Task-type X Feedback					
Hits	1	2.19	0.043	0.306	0.145
False Alarms	1	0.069	0.001	0.058	0.794
Sensitivity	1	0.205	0.004	0.073	0.653
Error	49				

*** $p < 0.001$

** $p < 0.01$

Given the unusual outcomes concerning the provision of feedback on the inspection task, the effects of this variable on inspection performance were reassessed after controlling for impulsivity. As suggested by Buss and Plomin (1974), high impulsives tend to act quickly without considering the consequences of their actions. This post hoc analysis was based on the rationale that low impulsives may have attended to the provision of feedback more intently than high impulsives. An ANCOVA was conducted to assess the effects of feedback on hit and false alarm rates on the inspection task using impulsivity as a covariate. Results suggest that differences in hit performance on the inspection task continue to exist between feedback conditions even after controlling for impulsivity. This was also repeated using boredom suscpetibility as a covariate with no significant changes in the already obtained outcomes.

Discussion

Predicting Vigilance Performance

The association between impulsivity and each performance measure (on the monitoring task) is consistent with a priori expectations. As discussed earlier, impulsivity appears to do better than extraversion in predicting vigilance performance, providing further evidence for the unreliability of extraversion in predicting vigilance performance. For example, the association between extraversion and hit performance is contrary to that consistently reported in previous research (see Koelega, 1992). The obtained positive association suggests that high extraversion is associated with better hit performance on this task, which is counter to that suggested by arousal theory which maintains that introverts outperform extraverts on this measure of performance (for example, Harkins & Geen, 1975; Eysenck, 1989).

The inconsistencies surrounding extraversion and vigilance were recently demonstrated in Koelega's (1989) meta-analysis, and were thought to result from either a failure to include extreme extraverts, the use of inappropriate statistical analyses, or inappropriate conclusions regarding extraversion and vigilance. However, present results suggest that a portion of these inconsistencies can be attributed to the measure of the construct itself. Extraversion is said to be made up of impulsivity and sociability (Eysenck, 1989). The NEOPI-R appears to measure extraversion primarily in terms of sociability, where some of the facets include warmth, gregariousness, and assertiveness (see Costa & McCrae, 1991). The differential effects of extraversion and its subscales (sociability and impulsivity) on certain behaviours have been previously demonstrated by Campbell (1983) and Revelle, Humphreys, Simon and Gilliland (1980), though not in the vigilance context. The latter group of researchers even suggested that impulsivity and sociability be looked at separately or in combination, depending on the nature of behavioural outcomes studied.

Even though impulsivity appears to bear more relevance to performance on vigilance tasks (Thackray, Jones & Touchstone, 1973), it has not been pursued as extensively as extraversion. The obtained medium effect sizes provide some support for the further exploration of impulsivity and vigilance. However, its usefulness in predicting vigilance performance appears to be limited to the monitoring task only, providing further insight into the inconsistencies surrounding past research on extraversion. Task-specificity may have played a role in the low effect sizes obtained in Koelega's meta-analysis, where different types of tasks may have been employed in the studies included in the meta-analysis. Similarly, the usefulness of boredom susceptibility and the digit symbol task (selective attention) may not only be specific to the inspection task (which was successive), but also specific to the performance measure (false alarms). Matthews and Holley (1993) provide some evidence for the task-specificity of resource-demanding predictors, where performance on tasks requiring attentional resources were found to predict perceptual sensitivity on three of four successive tasks, but only on one of four simultaneous tasks. The negative association between the digit symbol task and false alarms suggests that those scoring high on selective attention were less likely to commit false rejection errors on the inspection task. Similarly, those who scored high on the boredom susceptibility scale, were more likely to commit this same type of error, but only on the inspection task.

Overall, the predictability of vigilance performance from personality and cognitive factors appears to be dependent on the type of vigilance task and on the particular aspect of performance measured. With respect to the replication of previous findings (for example, Darr, 1998), it is unclear why selective attention failed to replicate itself as a predictor of hit performance on the monitoring task, accounting for 16% of the variance in hit performance in the previous study but zero variance in the present study. Task characteristics may have played some role in this outcome; for example, the monitoring task used in Darr (1998) employed the use of artificial signals, increasing the event-rate which may have contributed to an increased level of complexity and/or effort involved in that particular task. And as suggested in the preceding discussion, this cognitive (or resource-demanding) predictor could very well have differential associations with the task employed in the present study and the one employed in Darr (1998). This interpretation remains tentative at the present time, but future exploration may provide further insight.

Overall, the effect sizes of the predictor associations obtained in this study are promising; however the task-specificity and performance measure-specificity of these variables deter any attempts at developing universal selection criteria to be used across a range of vigilance tasks. To reiterate Koelega's (1989) discussion on the same matter, the serious practical need for predicting vigilance performance still prevails. However, adding to Koelega's perception that promising results in this area are more likely to be found by focussing on task characteristics (for example, information-processing), the combination of *operator* and *task* characteristics may account for vigilance performance better than either one alone.

Task-specific Performance Differences

Results provide support for most of the hypotheses regarding differential vigilance performance on the two tasks. The inspection task, according to vigilance literature on task parameters, was deemed to be more complex than the monitoring task. Data from the subjective measure confirms this categorization as 80% of the participants indicated that the inspection task was more complex. Significant performance differences between the monitoring and the inspection task are evident. Overall performance (target hits) was lower on the inspection task because, as hypothesized, the demands imposed by this particular task made it more difficult to detect critical signals than those imposed by the monitoring task.

Even though average performance decrement estimates were not found to differ significantly across the two tasks, a comparison of the beta-weights obtained when hit performance was regressed onto time suggests that the decline appears to be greater on the monitoring task than on the inspection task (see Figure 3). This finding is consistent with Loeb, Noonan, Ash and Holding's (1987) speculation that decrement on complex tasks may be arrested as participants learn strategies for dealing with the task. Even though there is no direct evidence for whether or not participants employed any strategies in their inspection of disk-drives, information from the subjective measure suggests that most participants found the inspection task required less effort as time on task passed by. This suggests that participants found it easier to discriminate between defective and non-defective disk-drives as time passed on. In addition, the number of times participants viewed the model differed across time sessions, where on average, the sample stimulus was viewed 21 times in the first session and gradually declined to five times in the last session. However, average hit performance across time on task does not provide strong evidence for this explanation as hit performance does not substantially improve with time on task (see Figure 3). Perhaps, over time participants became more confident of their ability to correctly recognize and retain a non-defective disk-drive (which contributed to their false-alarm rate) than to recognize defective ones (which contributed to their hit performance). A visual examination of the false alarm rate across time suggests that this explanation may be plausible, where average false-alarm rates dropped to zero in the last interval of the inspection task.

The remainder of Loeb et al's (1987) speculation of vigilance decrement on complex tasks concerned the efficient learning of critical signals such that the 'complex' task becomes monotonous enough to introduce an element of decrement. Anticipation of this effect led to the hypothesis that sensitivity shifts at the latter portion of the inspection task would be similar to that at the beginning of the monitoring task. However, this comparison was not possible as relatively poor hit performance on the inspection task suggests that participants were unable to efficiently detect critical signals at any time on the task. The inspection task was probably too difficult to master within the given time period, and a longer session may have perhaps permitted this comparison.

Future research may benefit from introducing either a longer training component or deferring testing until some criterion has been reached (for example, 80% correct responding, Mason & Redmon, 1992). Chaney and Teel (1967 found training to be beneficial in improving detection performance of machine-part inspectors. Pre-task training was suggested to be especially critical for tasks with excessive information-processing demands (Williams, 1986). Even though Williams succeeded in compensating for training by introducing a 50- minute long training component prior to a signal detection task, processing demands of the task (which employed a high event rate) may have been too high for optimal performance outcomes. Given the information processing demands of the inspection task used in this study, the effects of criterion referenced training on task performance may prove beneficial. After all, inspectors and radar operators in operational settings undergo extensive training and in some cases must meet acceptable standards prior to being placed on the job.

Even though hit performance was poorer on the inspection task, false-alarm rates on this task remained relatively low (when compared to that on the monitoring task). This finding appears to be consistent with Zenger and Fahle's (1997) discussion on error and false-alarm rates on visual search tasks of varying display sizes (making the task increasing difficult). These researchers found that missed-target errors (or hit performance in this case) were more common than false-alarms, with both performance measures increasing and decreasing, respectively, with increasing display sizes. In other words, as the size of the screen display increased, false alarm rates dropped, but there was an increase in missed targets (or failure to detect signals). Their explanation of this phenomenon borrows from the Guided Search (GS2) model (Chun & Wolfe, 1996), where larger displays (which require long search times in comparison to smaller displays) are scanned using the GS2 search termination rule, where the speed-accuracy trade off is greater. In other words, when a search has to be terminated without a target being found, the strategy is to guess, 97 percent of the time, that no target is present and to guess (in the remaining 3 percent of the cases) that a target is present. This model bears relevance to the inspection task because this task required the scanning of a stimulus configuration in search of a target (a missing element). Because of the time constraints imposed, it is possible that participants terminated their search using the GS2 guessing strategy, which led to a higher rate of missed targets (that is, an incorrect "no signal present" response), but a lower false-alarm rate (incorrect "signal present" response).

The non-linear distribution of false-alarm rates on the monitoring task can be best explained within the arousal frame-work. Stroh (1971) discussed the inconsistent findings between arousal and vigilance performance, suggesting that the relationship between arousal and time on task may be curvilinear rather than linear. A preliminary study reported in his discussion revealed that during a 1-hour vigilance task, participants' level of arousal (EEG) suddenly increased during the last 10-minutes of the task even though participants did not report any conscious awareness that the hour was up. Generally, high arousal is associated with better vigilance performance than that obtained when arousal is low. Drawing from Stroh's discussion, we can speculate that participants' level of arousal on the monitoring task may have increased during the latter portion of the task, resulting in an increased level of vigilance at that particular time on the task.

With respect to individual differences in performance across the two tasks, within-task performance variability on the inspection task is about five times the variability found within the monitoring task. This finding is consistent with previous studies (for example, Koelega et al.,1989; Parasuraman, 1976) and suggests that individual differences in performance is largely task-specific. In other words, participants performed similarly to each other across time on the monitoring task, but the same participants differed greatly from one another on the inspection task.

The Effects of Feedback

Quantitative and visual feedback, which was introduced in an attempt to improve performance on both task-types, produced mixed results. Average hit performance on both tasks appeared to be lower when feedback was present than when it was absent. However, the effect is more pronounced on the inspection task, where hit performance is almost 11 percent points higher when no feedback is given (see Figure 4). A visual examination of the average decrement plot (see Figure 5) suggests that feedback may have also disrupted performance over time on the inspection task, but not on the monitoring task. These findings are contrary to that suggested by previous research, but provide some insight into the future application of feedback in vigilance tasks.

Balcazar, Hopkins, and Suarez (1985) critically reviewed performance feedback in 126 organizations to find that, contrary to previous literature, feedback is *not* uniformly effective. Their objective analysis revealed that the effects of feedback may be improved by introducing functional consequences. In other words, combining tangible positive rewards with performance feedback is more likely to have a reinforcing effect on the desired behaviour than when feedback is provided alone. Nachreiner (1977) described how the experimental vigilance paradigm has very few intrinsic motivational properties, which may contribute to poor performance. Functional consequences may, therefore, provide the necessary extrinsic motivation, leading to improved performance. As feedback was not paired with any external consequence in the present study, this approach is worth further examination and may produce fruitful effects on vigilance performance.

Although the negative effects of feedback on monitoring performance, in the present study, can be described as minimal (that is, it produced only a 2-3 point difference in hit performance), its effects on inspection performance raises some concern. Participants appeared to perform poorer overall, in terms of total hit performance and performance over time on task. The physical dimensions of feedback (size, location, etc.) may have played a role in the obtained findings. Perhaps, the presence of feedback (an additional stimulus) on the already complex visual display of the inspection task disrupted participants' concentration on the task. Evidence for this speculation is provided by Craig (1981) who demonstrated that signal detection can be impaired in the mere presence of an additional signal (or stimulus) even when this signal does not have to be detected. In essence, the physical presence of the feedback icon on the screen display of the inspection task used in this study may have interfered with participants' inspection performance.

Feedback, operationalized as a quantitative running percent score, may have further distracted participants' concentration on the inspection task. Because an increase in the percent score reflected a correct response while a decrease reflected an incorrect response, participants had to be aware of the running percent score at all times in order to know whether a particular response was correct or wrong. In other words, the presentation of feedback in this manner may have reduced its effectiveness in improving performance. Perhaps, the provision of qualitative feedback (having the words "correct" or "wrong" appear after a response) may have been more effective. Furthermore, results showed that feedback disrupted performance on the inspection task, suggesting that presentation increased the complexity of an already complex task.

Findings suggest that, in addition to examining the motivational properties of performance feedback, it may be necessary to carefully consider the 'dimensionality' (that is, size, location, presentation) of feedback to be administered as well. The latter may, however, be of concern only on tasks involving complex visual displays. Therefore if feedback is to be optimally used in various settings, especially in the vigilance context, it is important to explore its characteristics and the surrounding conditions within which it is expected to produce desired effects.

Limitations of the Present Study

The examination of vigilance performance in this study was confined to task characteristics. If future vigilance research is to improve its generalizability, other dimensions (environmental and temporal characteristics) described in Davies and Parasuraman's (1982) matrix of operational relevance must also be included in our examination. In industrial settings, for example, individuals rarely work in isolation. The effect of working in groups and in the presence of realistic noise events must be considered in future experimentation on vigilance. As discussed earlier, the effects of training must also be explored. Mackie (1977) discussed how optimal performance can be elicited from individuals who are better able to identify with the given task (that is, those who do the same kind of work). Data from a previous study (Garner, 1998), which compared monitoring performance of professional radar operators to a student sample, confirms Mackie's suggestion. The professional group outperformed the student sample on overall level of detection and on performance over time. The use of a convenient sample is likely to have restricted the range of scores on the predictor variables, limiting the magnitude of true associations between variables. Methodological issues concerning the delivery of feedback on the inspection task may have played a role in the unusual outcomes concerning feedback and vigilance performance. However, suggestions for improving the strength of this manipulation as discussed previously may be incorportated in future research.

Practical Applications and Implications

One of the main purposes of this study was to determine the extent to which findings from past vigilance research could be used in examining or even predicting vigilance performance across two different task-types. The obtained results are promising, in that the main hypotheses concerning differential performance across the two tasks are supported. Consistent with expectations, overall vigilance performance was better on the simple vigilance (monitoring) task in comparison to the complex one (inspection). Findings also suggest that vigilance decrement, a much-debated phenomenon, is not limited to simple vigilance tasks. Performance on complex tasks can also be expected to decline, though its effects may not be as pronounced as on simple tasks. This finding is consistent with that obtained by Pigeau et al. (1995), where performance decrement on an operational aircraft surveillance task (assumed to be more complex than typical vigilance tasks in laboratory studies) was found to be less strong than that reported in laboratory findings.

The impetus for comparing vigilance performance across the two task-types was to draw inferences that relate to operational tasks, which literature describes as being more complex than those used in lab settings. Using Mackie's (1977) matrix of operational relevance, the inspection task used in this study was previously described as having more operational relevance than the monitoring task. As the quality control director confirmed , the inspection task is not unlike some of the tasks performed by quality control inspectors in current industrial settings. Therefore, if the inspection task can be described as being similar to some of tasks in the real-world, then the obtained results are not surprising. In other words, it wouldn't be unreasonable to expect similar performance outcomes (as obtained on the inspection task) if the convenient sample of participants used in this study were to perform a vigilance task typical of current industrial settings.

The present study was designed taking into consideration some of the criticisms of vigilance research and heeding to the directions offered by experts in this area. Overall, findings provide support for Weiner's (1987) description of vigilance research to date as being "well-done", in that, we *can* depend on past findings in expecting certain performance outcomes on related vigilance tasks. The comparison of feedback (an external manipulation) across the two different vigilance tasks, an approach suggested by Mackie (1987), provides some insight into improving the effects of this variable on the two tasks. From an operational perspective, cross-task comparisons of certain manipulations like this one may prove beneficial when designing future performance systems.

Proceeding with the advice of vigilance experts, the next step would be to develop counter measures aimed at improving particular aspects of performance on both the monitoring and the inspection task used in this study. For example, measures aimed at reducing vigilance decrement on the monitoring task and those aimed at improving overall detection on the inspection task may provide focus for future studies. Cues to "how" and "what" these measures may comprise are manifested through findings in the present study. For example, decrement on the monitoring task may be reduced by introducing an element of complexity (perhaps, through artificial signal injection which is likely to increase one task parameter, event-rate). Similarly, the detection of signals on the inspection task may be improved through criterion referenced training. The determination of optimal physical and motivational properties of feedback on both tasks may also be included as a measure aimed at improving performance on these two tasks. Exploration of countermeasures such as these and the continued use of more operationally relevant simulations is likely to improve the current vigilance research situation in the long run. Ultimately, our ideal would be the examination of vigilance in an applied setting. While that remains our long-term goal, the only viable alternative at the present time is to *simulate*.

References

Adams, J.A., Humes, J.M., & Sieveking, N.A.(1963). Monitoring of complex visual displays: V. effects of repeated sessions and heavy visual load on human vigilance. Human Factors, 29, 737-340.

Adams, J.A. (1987). Criticisms of vigilance research: A discussion. Human Factors, 29(6), 737-740.

Badalamente, R.V. (1969). A behavioral analysis of an assembly line inspection task. Human Factors, 11(4), 339-352.

Baker, C.H. (1960). Maintaining the level of vigilance by means of artificial signals. Journal of Applied Psychology, 44(5), 336-338.

Balcazar, F. Hopkins, B.L. & Suarez, Y. (1986). A critical, objective review of performance feedback. Journal of Organizational Behavior Management, 7(3/4), 65- 87.

Buckner, D.N., Harabedian, A., & McGrath, J.J. (1960). A study of individual differences in vigilance performance. In D.R. Davies & R. Parasuraman, Psychology of Vigilance. New York: Academic Press.

Buss, A.H. & Plomin, R. (1975). A Temperament Theory of Personality Development. New York: John Wiley & Sons.

Campbell, J.B. (1983). Differential relationships of extraversion, impulsivity, and sociability to study habits. Journal Research in Personality, 17, 308-314.

Chaney, F. B & Teel, K.S. (1967). Improving inspector performance through training and visual aids. Journal of Applied Psychology, 51(4), 311-315.

Chun, M.M. & Wolfe, J.M. (1996). Just say no: How are visual searches terminated when there is no target present? Cognitive Psychology,30, 39-78.

Craig, A. (1981). Monitoring for one kind of signal in the presence of another: The effects of signal mix on detectability. Human Factors, 23(2), 191-197.

Craig, A. (1984). Human Engineering: the control of vigilance. In J.S. Warm (Ed.). Sustained Attention in Human Performance. New York: John Wiley & Sons Ltd.

Costa, P. T., Jr. & McCrae, R.R. (1991). Professional Manual: Revised NEO Personality Inventory. Florida: Psychological Assessment Resources, Inc.

Darr, W.A. (1998). Predicting performance on Monitoring tasks. Poster presented at the 13th Annual SMU Psychology Conference, Saint Mary's University, Halifax, N.S.

Davies, D.R. & Parasuraman, R. (1982). The Psychology of Vigilance. New York: Academic Press.

Eysenck, M.W. (1989). Individual differences in vigilance performance. In A. Coblentz (Ed.). Vigilance and Performance in Automatized Systems. Dordrecht: Kluwer Academic Publishers.

Garner, A. (1998). Comparison of performance across professional and paid and unpaid student samples on a vigilance task. Paper presented at the 13th Annual Psychology Conference, Saint Mary's University, Halifax, March.

Gerbing, D.W., Ahadi, S.A. & Patton, J.H. (1987). Toward a conceptualization of impulsivity: Components across the behavioural and self-report domains. Multivariate Behavioural Research,22, 357-379.

Grant, K. (1997). The effects of artificial signal injection, performance feedback, and personality on vigilance performance. Unpublished Honours thesis, Saint Mary's University, Halifax, N.S.

Grier, J.B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas.

Psychological Bulletin,75, 424-429.

Groth-Marnat, G. (1990). Handbook of Psychological Assessment (2nd ed.). New York: John Wiley & Sons.

Harkins, S. & Geen, R.G. (1975). Discriminability and criterion differences between extraverts and introverts during vigilance. Journal of Research in Personality, 9, 335-340.

Harris, D. (1966). Effect of equipment complexity on inspection performance. Journal of Applied Psychology, 50(3), 236-237.

Harris, D. (1969). The nature of industrial inspection. Human Factors, 11, 139-148.

Johnston, W.A., Howell, W.C., & Goldstein, I.L. (1966). Human vigilance as a function of signal frequency and stimulus density. Journal of Experimental Psychology, 72(5), 736-743.

Kahneman, D. (1973). Attention and Effort. New Jersey: Prentice-Hall Inc.

Koelega, H.S., Brinkman, J.A., Hendriks, L., & Verbaten, M.N. (1989). Processing demands, effort, and individual differences in four different vigilance tasks. Human Factors, 31(1), 45-62.

Lanzetta, T.M, Dember, W.N., Warm, J.S., & Berch, D.B. (1987). Effects of task type and stimulus heterogeneity on the event rate function in sustained attention. Human Factors, 29(6), 625-633.

Levine, J.M., Romashko, T. & Fleishman, E.A. (1973). Evaluation of an abilities classification system for integrating and generalizing human performance research findings: An application to vigilance tasks. Journal of Applied Psychology, 58(2), 149-157.

Loeb, M., Noonan, T.K., Ash, D.W. & Holding (1987). Limitations of the cognitive vigilance increment. Human factors, 29(6), 661-674.

Mackie, R.R. (1977). Introduction. . In R.R. Mackie (Ed.). Vigilance: Theory, Operational Performance, and Physiological Correlates (vol. 3, pp. 665-678). New York: Plenum Press.

Mackie, R.R. (1987). Vigilance Research- Are we ready for counter measures? Human Factors, 29(6), 707-723.

Mackie, R.R., Wylie, C.D., & Smith, M.J. (1994). Countering loss of vigilance in sonar watchstanding using signal injection and performance feedback. Ergonomics, 37(7), 1157-1184.

Mackworth, N.H. (1948). The breakdown of vigilance during prolonged visual search. Quarterly Journal of Experimental Psychology, 1, 6-21.

Mason, M.A. & Redmon, W.K. (1992). Effects of immediate versus delayed feedback on the error detection accuracy in a quality control simulation. Journal of Organizational Behavior Management, 13(1), 49-83.

Matthews, G. & Holley, P.J. (1993). Cognitive predictors of vigilance. Human Factors, 35(1), 3-24.

Methot, L. (under review). Monitoring for one type of critical stimulus event in the presence of a second event: The role of performance feedback.

Methot, L. & Huitema, B.E. (1998). Effects of signal probability on individual differences in vigilance. Human Factors, 40(1), 102-110.

Nachreiner, F. (1977). Experiments on the validity of vigilance experiments. In R.R. Mackie (Ed.). Vigilance: Theory, Operational Performance, and Physiological Correlates (vol. 3, pp. 665-678). New York: Plenum Press.

Peterson, N. (1982). The Behaviour Analyst. Kalamazoo: Western Michigan University.

Pigeau, R.A., Angus, R.G., O'Neill, P., & Mack, I. (1995). Vigilance latencies to aircraft detections among NORAD surveillance operators. Human Factors, 37(3), 622-634.

Parasuraman, R. (1976). Consistency of individual differences in human vigilance performance: An abilities classification analysis. Journal of Applied Pscyhology,61(4), 486-492.

Parasuraman, R. (1985). Sustained Attention: A multifactorial approach. In M.I. Posner & O.S.M. Marin (Eds.). Attention and Performance. New Jersey: Erlbaum.

Parasuraman, R. (1987). Human computer monitoring. Human factors, 29, 695-706.

Parasuraman, R. & Mouloua, M. (1987). Interaction of signal discriminability and task type in vigilance decrement. Perception & Psychophysics, 41(1), 17-22.

Poulton, E.C. (1973). The effects of fatigue upon inspection work. In D.R. Davies and R. Parasuarman. Psychology of Vigilance. New York: Harcourt Brace Jovanovich.

Prue, D.M. & Fairbank, J.A. (1981). Performance feedback in organizational behavior management: a review. Journal of Organizational Behavior Management, 3, 1-16.

Revelle, W., Humphreys, M.S. & Gilliland, K. (1980). The interactive effects of personality, time of day, and caffeine: A test of the arousal model. Journal of Experimental Psychology: General, 109, 1-31.

See, J.E., Howe, S.R., Warm, J.S., & Dember, W.N. (1995). Meta-analysis of the sensitivity decrement in vigilance. Psychological Bulletin, 117(2), 230-249.

Smith, R.I. & Lucaccini (1969). Vigilance Research: Its application to industrial problems. Human Factors, 11(2), 149-156.

Stanislaw, H. (1995). Effect of type of task and number of inspectors on performance of an industrial inspection-type task. Human Factors, 37(1), 182-192.

Stroh, C. M. (1971). Vigilance Performance: The Problem of Sustained Attention. New York: Pergamon Press.

Tanner, Jr., W.P. & Swets, J.A. (1954). A decision-making theory of visual detection.

Psychological Review, 61(6), 401-409.

Thackray, R.I., Jones, K.N., & Touchstone, R.M. (1974). Personality and physiological correlates of performance decrement on a monotonous task requiring sustained attention. British Journal of Psychology, 65(3), 351-358.

Thackray, R.I. & Touchstone, M. (1991). Effects of monitoring under high and low taskload on detection of flashing and coloured radar targets. Ergonomics, 34(8), 1065-1081.

Weiner, E.L. (1987). Application of vigilance research: Rare, medium, or well done? Human Factors, 29(6), 725-736.

Wechsler, D. (1981).WAIS-R Manual: Wechsler Adult Intelligence Scale-Revised. New York: Psychological Corporation.

Williams, P.S. (1986). Processing demands, training, and the vigilance decrement. Human Factors, 28 (5), 567-579.

Windle, M. (1989). Temperament and Personality: An exploratory interinventory study of the DOTS-R, EASI-II, and EPI. Journal of Personality Assessment, 53(3), 487-501.

Zenger, B. & Fahle, M. (1997). Missed targets are more frequent than false alarms: A model for error rates in visual search. Journal of Experimental Psychology: Human Perception and Performance, 27(6), 1783-1791.

Zuckerman, M., Eysenck, S. & Eysenck, H.J. (1978). Sensation Seeking in England and America: Cross-cultural, age, and sex comparisons. Journal of Consulting and Clinical Psychology, 46 (1), 139-149.

Zuckerman, M. (1995). Abbreviated manual with Scoring keys and Forms for Form V of the Sensation Seeking Scale. Newark: University of Delaware.

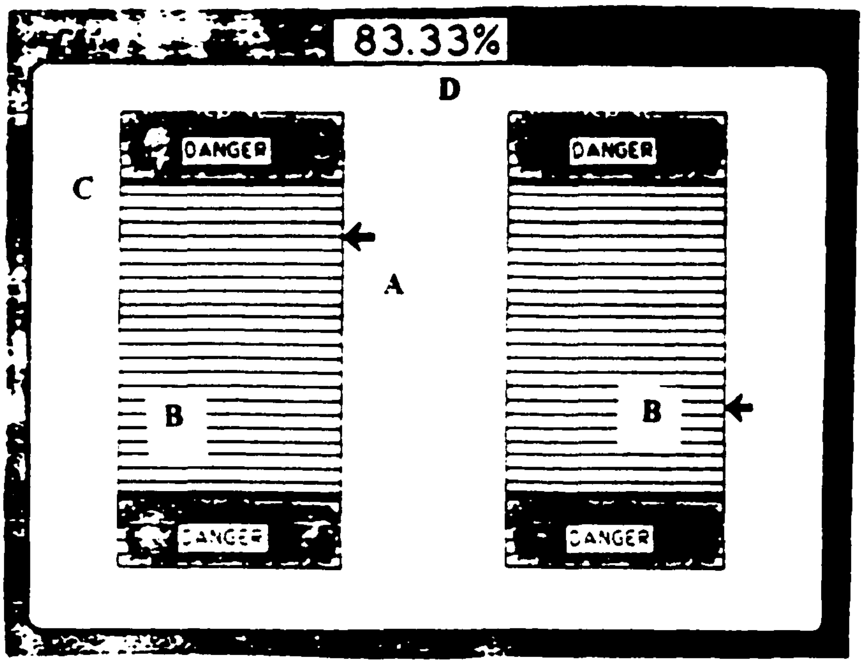
Appendix A

Typical Screen displays viewed by participants during the Monitoring and Inspection Task

MONITORING

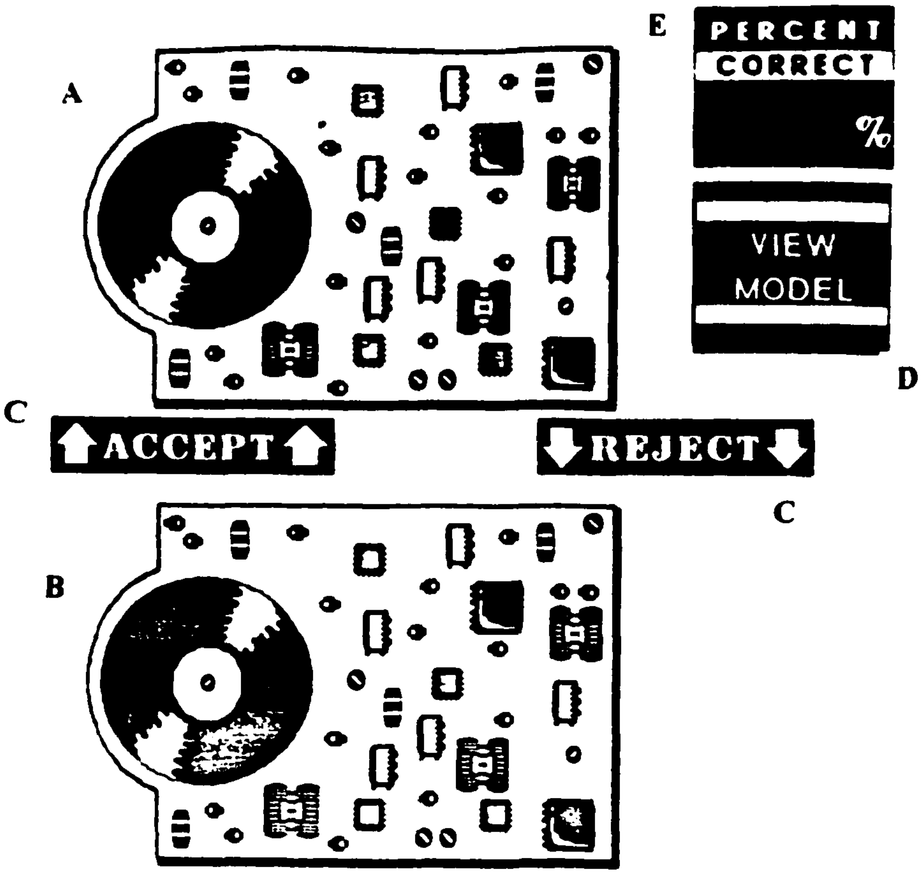
LEGEND

A = signal
B = noise zone
C = meter gauge
D = feedback indicator



INSPECTION

A = sample disk drive
B = model stimulus
C = response icons
D = view model icon
E = feedback indicator



Appendix B

Subjective Measure

In your opinion, was one task more complex than the other? Please check-mark one of the following:

- ☐ the Monitoring task was more complex
- ☐ the Inspection task was more complex
- ☐ both tasks were equally complex
- ☐ none of the tasks were complex

Did you exert more effort at the beginning or the end of each task? Please answer the following set of questions:

The Monitoring task

- ☐ required less effort as time passed on
- ☐ required more effort as time passed on
- ☐ did not require any more effort at any particular time on the task

The Inspection task

- ☐ required less effort as time passed on
- ☐ required more effort as time passed on
- ☐ did not require any more effort at any particular time on the task