Team Performance on Monitoring Tasks: An Examination of Decision Errors in Contexts Requiring Sustained Attention

John R. Hollenbeck, Daniel R. Ilgen, Dale B. Tuttle, and Douglas J. Sego[1]

Department of Management
Michigan State University
East Lansing, Michigan 48824

Abstract

The purpose of this study was to examine individual and team performance in situations requiring sustained attention. Eighty research participants arrayed in 20 four-person teams worked for two separate 3-hour sessions on a naval command and control simulation. The results of the study replicate the past literature on the vigilance decrement and extend the literature by documenting a post-critical signal decrement as well. The study also shows that problems in the area of sustained attention generalize from situations involving individual decision makers to contexts where decisions are made by teams.

1. Introduction

Predicting and enhancing the level of human performance has been one of the major goals of research in applied psychology. Despite this fact, [Campbell, 1991] recently noted that success in this area has been limited because researchers typically approach the performance problem as if performance on any job is a singular, unidimensional construct. Future progress in this area, according to Campbell, hinges on shifting the focus from the overall job level (which is multi-dimensional) to the job component level (which is unidimensional). Such a focus could lead to a more meaningful accumulation of knowledge about important components of job performance that span multiple jobs and occupations.

One important component of performance in many jobs is that of sustained attention. We will formally define sustained attention as a task demand that forces the job incumbent to remain alert and actively engaged in a sequential and repetitive information processing task over a prolonged period of time. In a context requiring sustained attention, events occurring later in the sequence may or may not be dependent on earlier events, but the job incumbents potential awareness of the entire sequence may influence that person's reactions to later events.

This particular aspect of human performance has always been present in many types of jobs, however, increased levels of automation in modern work organizations have made this a particularly salient issue. As noted in the [Human Capital Initiative, 1993], published by the American Psychological Society, "one of the most striking psychological aspects of high tech jobs is that motor, perceptual, and even some decision-making requirements are being amended by automated devices and 'intelligent' computer programs" (p. 13).

In general, automation relieves the job incumbent from many of the routine, but controlling activities that were previously necessary with less sophisticated systems. Instead of being placed in the role of a controller, the person is placed in the role of a passive observer of dials, video screens, and other sources of information [Kessel & Wickens, 1982]. The individual's primary responsibility is to monitor these instruments in order to detect critical situations and to make decisions or initiate actions should critical situations arise. A large number of jobs require this kind of sustained attention, including airport security, radar/sonar operations, nuclear power plant operations, air/sea

[1]This research was supported, in part, by Grant No. N00014-90-J-1788 from the Office of Naval Research as part of the technical base research for the Cognitive and Neural Sciences Division. Although support for this research is gratefully acknowledged, the ideas expressed within are those of the authors and not necessarily those of the agency.

navigation, industrial quality control, prison guards, police surveillance, long-haul trucking, railway operations, air traffic control, anesthesiology, and nursing.

The increasing role that passive monitoring and surveillance plays in contemporary job descriptions has been viewed with alarm by those familiar with the literature on vigilance. Over forty years ago, researchers noted that human performance deteriorates rapidly on routine monitoring tasks [Mackworth, 1948]. After roughly one thousand studies of this phenomenon, researchers in this area have concluded that this vigilance decrement is "about as dependable a result as one will ever see in human experimentation" [Wiener, 1987, p. 729-730]. Vigilance decrements have been cited as factors contributing to many real-world disasters such as the KAL terrorist bombing (attributed to inattentive security agents), the Three Mile Island nuclear accident, the grounding of the Exxon Valdez, the destruction of the U.S.S. Stark by an Iraqui attack plane, the Bhopal chemical disaster, and several commercial aircraft crashes [National Transportation Safety Board, 1984; 1986].

The purpose of this research is to study human performance in a context requiring sustained attention. Vigilance plays a major role in such situations, but it is only part of the process. Sustained attention occurs across the whole task sequence including the time before, during and after the appearance of the critical event typically studied in vigilance research. The current study will explore implications of performance in contexts requiring sustained attention for both individuals and teams. Given the major role that vigilance plays in contexts requiring sustained attention, the literature on the vigilance decrement will be discussed below.

1. 1 Research on Vigilance

Systematic controlled experimentation in the area of vigilance first began with a series of studies by [Mackworth, 1948]. Research participants in Mackworth's original studies sat at a simulated radar display and monitored the movement of a black pointer that shifted .3 inches every second. Infrequently, and at irregular intervals, the pointer jumped .6 inches. This was the critical signal and the subjects' task was to press a button every time they detected such a double length jump. Mackworth noted that the incidence of missing the critical signal increased sharply from the first to second half hour on the task. The rate of missing the critical signal then increased only minimally after this point for the remainder of a two hour vigil. This pattern of performance deterioration has since been replicated using different types of tasks (visual, auditory, cutaneous). Taken as a whole, these studies suggest that the vigilance decrement is typically complete within 35 minutes after the initiation of the vigil [Teichner, 1974].

Since these original studies, there has been wide variability in tasks studied under the heading of vigilance, and thus it is important to stress the common features that seem to comprise a vigilance task. First, the task is prolonged and continuous, often lasting for hours. Second, the critical signal, that is, the stimulus to be detected, is usually not compelling, but it is clearly perceptible when the observer is alerted to it. Third, the critical signal occurs infrequently, irregularly and without warning. Finally, responses on the part of the observer are unrelated to the probability of appearance of the critical signal. Although not all vigilance tasks conform completely to these specifications, these dimensions capture most vigilance tasks. The similarity between traditional vigilance tasks and any task requiring sustained attention are obvious.

As in other areas of psychology, recent advances in cognitive psychology and information processing have influenced vigilance researchers [Fisk, Ackerman & Schneider, 1987]. Application of a cognitive perspective to the problem of vigilance highlights the need to distinguish between automatic and controlled modes of information processing. Information processing in automatic mode is characterized as fast, effortless, and proceduralized. Processing information in this manner generally does not require conscious attention, and therefore, automatic processing allows for simultaneous and parallel operation of other tasks.

Information processing in controlled mode, on the other hand, is characterized as slow, effortful, and non-standardized. This type of information processing occurs serially and greatly reduces one's ability to perform other information processing tasks simultaneously [Schneider & Shiffren, 1977; Shiffren & Schneider, 1977].

From an information processing perspective, the problem created by vigilance tasks is that the low rate of critical events lulls the observers into a state of automatic information processing. In this state, attention may stray from the task, and observers become ill-prepared for the arrival of the critical signal. Thus, based upon the literature on vigilance and information processing theory, our first hypothesis is that individual decision makers performing in a context requiring sustained attention will exhibit a performance failure on the critical event.

1. 2 Decision Heuristics in Contexts Requiring Sustained Attention

Whereas the vigilance literature speaks to performance in sustained attention contexts up through the occurrence of the performance decrement on the critical event, it has less to say about what will happen following the critical event. If the job incumbent fails to respond appropriately to critical event, this failure rarely goes unnoticed by the job incumbent or those that surround this person, and the awareness of this failure could have implications for the incumbent's ability to manage subsequent non-critical signals. In sustained attention contexts, performance after the critical event is just as important as performance on the critical event, and therefore the dynamic effect of past errors on future decisions cannot be discounted.

The cognitive perspective mentioned above may also be helpful for understanding performance under sustained attention conditions following failure or some non-routine event. In particular, cognitive perspectives highlight the role of decision heuristics that enter into the judgment process, as well as dynamic influences of past decision outcomes on current information processing e.g., [Hollenbeck, Ilgen, Phillips & Hedlund, 1994]. Specifically, one well-known finding in the literature on decision heuristics that seems relevant when analyzing vigilance decrements as judgment errors is the availability bias. The availability bias occurs when events that are easy to recall have a greater impact on decisions than is warranted given their objective probability or ecological validity [Tversky & Kahneman, 1981]. Given the structure of human memory [Anderson, 1983], highly salient events (that may be salient precisely because they are rare) and events that have occurred more recently are easier to recover from memory than events characterized otherwise. Thus, in contexts requiring sustained attention, the impact of the critical signal (due to its recency and salience) may have serious negative effects on one's ability to manage subsequent non-critical signals. Although vigilance research has occasionally examined false alarm rates in general, it has not specifically addressed the false alarm rate for signals that immediately follow the critical signal.

Indeed, the negative effect of the critical signal on the next non-critical signal may be equal to or greater than the negative effect of the background signals on the critical signal. One of the major purposes of this study is to examine this possibility. Based upon the literature on decision making heuristics, our second hypothesis is that individual decision makers performing in a context requiring sustained attention will also exhibit a performance failure on the post-critical event.

1. 3 Sustained Attention in Teams

A second major purpose of this study is to examine teams performing under conditions of sustained attention. Many individuals occupying jobs that involve sustained attention work closely with others in groups or teams (e.g., airline crews, airport security staffs, anesthesiologists, command and control centers, police surveillance units, nurses, etc.). Indeed, teams are often created to help make these otherwise boring jobs more interesting. Moreover, these teams are typically configured as hierarchical decision-making teams with distributed expertise among staff members. That is, there are status differences among team members (one member is a leader), as well as differences in areas of expertise among staff members that make them non-interchangeable.

Given that many jobs that incorporate a sustained attention component take place within teams, it is regrettable that the vast amount of literature on vigilance contains, to our knowledge, no studies where individuals work in groups or teams. Thus, it is difficult to draw hypotheses directly from the vigilance literature when decision making teams are the unit of analysis. From other literatures, two competing team effects seem plausible.

First, placing a task requiring sustained attention in a team context may reduce the likelihood of error in the case of critical events. In terms of information processing, whereas the odds of missing the critical signal may be high for any one individual (e.g., .75), the combined probability that all members of a four-person team would miss the critical signal is not nearly so alarming ($.75^4 = .32$). In the group decision making literature, this is referred to as an assembly bonus effect [Collins and Guetzkow, 1964, Michaelson, Watson, and Black, 1989].

On the other hand, simply placing the task within a team context may have little impact on error rates for several reasons. First, it is clear from the literature on group decision making that even though one member of the group identifies a flaw in a group's decision, it does not necessarily follow that the group's decision will be altered. Self-censorship and other forms of consensus-seeking behavior on the part of the group could vitiate any beneficial effects attributable to a lone dissenter [Janis, 1982]. Indeed, empirical comparisons between real decision making teams and computer simulations of group behavior indicate that solutions from computer models that reflect a "truth wins" process rarely match the actual decisions arrived at by real groups [Laughlin, 1980; Tinsdale & Larson, 1992]. Second, the presence of others could distract attention from the task, especially since the repetitive nature of tasks requiring sustained attention tends to bore human operators. Third, working as part of a group might also promote social loafing [Latane, Williams, & Harkins, 1979] on the part of some team members who fail to put in the effort necessary to process each signal in a controlled fashion.

Thus, there are clearly arguments on both sides of the issue of whether placing a task requiring sustained attention in a group context will eliminate performance failures. While recognizing that there are arguments on both sides of his issue, we nonetheless believe that the literature on group dynamics would suggest that the problems associated with human performance in contexts requiring sustained attention will not be eliminated in team contexts. Thus, <u>our third hypothesis is that decision making teams performing in a context requiring sustained attention will exhibit a performance failures on both the critical and post-critical target</u>.

2. Method

2. 1 Subjects

Research participants were 80 male undergraduate students at a large midwestern university who were arrayed into 20 four-member teams. All were paid an hourly rate for their participation ($5.00), and worked 6 three-hour sessions over a period of six weeks. In addition to this, all had the opportunity to earn additional bonus money contingent upon their level of performance. The top performing team earned an additional $80.00, the next best team earned $40.00, and the third best team earned $20.00. Only two of these sessions (the second and fifth) dealt with performance in contexts requiring sustained attention, however, and the four remaining sessions were dedicated to a different study reported in [Hollenbeck, Ilgen, Sego, Hedlund, Major and Phillips (in press)]. No data reported here overlapped with the data reported in Hollenbeck et al. (in press).

2. 2 Task Overview

Research participants worked at a team-based version of the TIDE[2] (Team Interactive Decision Exercise for Teams Incorporating Distributed Expertise) simulation task. A brief description of this task is provided below. The interested reader is referred to [Hollenbeck, Sego, Ilgen, and Major, 1991] for a more complete description of this task, as well as documentation on this software and how it can be used.

In general, TIDE[2] is a software program for a decision task simulation that presents participants with a computerized multiple cue probability learning task, typically under conditions of feedforward instructions [Stevenson, Busemeyer, & Naylor, 1991]. That is, research participants are asked to make decisions about the state of some object (e.g., the desirability of a job candidate, the value of a particular piece of merchandise, the severity of a specific injury) based upon an evaluation of information provided on various cues. Prior to being presented with the object, research participants are told what characteristics of the object are relevant for this decision, and how the characteristics should be combined and weighed to reach a decision regarding the object.

In this study, $TIDE^2$ was programmed to simulate a naval command and control scenario. Participants were placed in four-person teams with four distinct roles. These roles consisted of the commanding officer (CO) on an aircraft carrier and three staff members who were monitoring the airspace from one of three locations -- an AWACS aircraft, an Aegis cruiser, and a land-based coastal air defense unit. Each of the individuals occupying staff positions were trained to possess unique areas of expertise related to the air patrol task.

The team's task was to monitor the airspace surrounding the Carrier for a two hour block of time. To do this, each team member was seated at a computer and observed a monitor. When an aircraft came into their airspace, a signal on the monitors of all four team members indicated its presence and the amount of time remaining for the team to reach a decision on how to respond to the incoming aircraft. For each aircraft (i.e., trial), each team member needed to gather information about the aircraft on the attributes (e.g., speed, direction, angle, range, size, etc.) that fell within that person's area of expertise, and then arrive at a judgment regarding the appropriate response to make toward the incoming aircraft. The recommendation was then sent to the leader, who was to decide how the team should respond.

2. 2 A Training on Task

There were two primary components in the training given to research participants. First, people learned the mechanics of gathering and sharing information. Each team member's work station was networked to all other team members, as well as a server that presented information on incoming aircraft. All collecting of information about an aircraft and communicating with fellow team members went on through the network. Participants were trained to perform all the operations necessary for obtaining or sharing information. A countdown clock on the screen indicated the time remaining in which a decision had to be made, and a warning signal (i.e., a high pitched beep) was activated when 30 seconds remained. Staff members were trained to get their recommendations in as soon as this warning indicator started to beep. The recommendations were chosen from a seven-valued continuum that ranged from high to low aggressiveness, including responses of defend, lock-on, ready, warn, monitor, review or ignore. After receiving all the recommendations from their staff, the leader then considered these opinions, along with any raw data in the leader's possession, and then entered the decision for the team using the same seven-valued scale. Teams had, on average, 210 seconds to respond to each incoming aircraft.

2. 2 B Training on Roles

The second part of the training dealt with each person learning their role. Each team member had unique expertise. That expertise was taught in the training session and came in three forms; (a) the ability to measure attributes of the aircraft, (b) the ability to translate raw data on aircraft attributes into judgments about the level of threat posed by the aircraft and, (c) knowledge of rules. For example, although all team members knew that aircraft traveling fast were more threatening than slow-moving aircraft, only two people in the team could actually measure speed (i.e., Attribute #1), and make an inference about the aircraft on this dimension given raw data in miles per hour (mph) units.

In addition to learning about aircraft attributes, each team member learned one of the rules about how attributes combined to affect threat. For example, one member memorized how speed and direction go together. Thus, at least one member of every team was an expert on one of the combination rules.

One last feature about the team's task involved interdependency. Although, four of the five rules for determining actual threat involved combinations of attributes, no one team member could measure both of the attributes necessary to satisfy any combination rule. That is, the team member who knew that speed and direction combined to affect threat was not able to measure both speed and direction. Information on one of the two attributes had to be obtained from some other team member. Thus, team members were interdependent.

2. 2 C Team Performance

The team decision was entered into the computer by the team leader. This decision was then compared to the correct decision which was based upon a linear combination that mirrored the set of rules given to participants. For example, focusing on only one rule (i.e., the "speed/direction" rule) the potential threat was determined by the formula:

$$Y_{t/sd} = (0.0 * X_s) + (0.0 * X_d) + (1.0 * X_s * X_d)$$

where,
$Y_{t/sd}$ = level of threat given speed and direction,
X_s = speed, and
X_d = direction

Note that the rule is established so that neither the main effect for speed nor for direction influenced threat. Only when both speed and direction were both known was it possible to draw an inference about the level of threat posed by the aircraft on these dimensions. Finally, although each cue was described in units appropriate to the cue (e.g. feet for altitude, degrees of angle of flight), all units were converted to a common metric before their values (X_is) were entered in the equation. The common metric was that of the threat level and was simply created by trichotimizing each dimension and assigning a threat value of 0.0 for the bottom third, 1.0 for the middle third and 2.0 for the top third of the range. For each aircraft, the combination of the speed/direction rule with the other four rules determined the overall level of threat associated with the aircraft.

The accuracy of the team's decision was determined by comparing the team's decision to the correct level of threat determined by the overall threat equation. The accuracy of the individual staff members was assessed by comparing that person's recommendation to the correct level determined by the overall equation. Feedback on the accuracy of the team's performance, as well as the individual staff members was presented immediately following the point at which the team leader entered the team's decision.

If the decision or recommendation matched the correct decision perfectly, this was termed a "hit." Being off by one point was referred to as a "near miss" (e.g., the operator said "defend" when "lock-on was correct). Being off by two points was referred to as "miss," being three points off resulted in an "incident," and being four or more points off resulted in a "disaster." The feedback screen also tracked the team's performance history in terms of the number of different kinds of outcomes (hits, misses, etc.,). It also provided, a projection of what the team's total set of outcomes would be at the end of the experiment if they continued to perform at the same level. The feedback remained on the screen for 15 seconds, after which, a new aircraft entered the region, and the next trial began.

3. Research Design

3. 1 Signal Type

This study used a within subjects research design, where our interest was in performance on three different types of signals (i.e., incoming aircraft): (a) the critical signal, (b) the signal that immediately preceded the critical signal (i.e., the pre-critical signal), and the signal immediately following the critical signal (i.e., the post-critical signal).

Table 1 shows the structure of the signals (i.e., incoming aircraft) that were presented to research participants in the two vigilance sessions. Each session included 18 trials, although our interest was primarily centered around Trials 9, 10, and 11. The other trials, labeled background trials, presented aircraft that were

Table 1

The Sequence of Aircraft Presented within Sessions

Aircraft Number	Signal Type	Correct Decision in Session 1[a]	Correct Decision in Session 2[b]
1)	Background	Defend/Lock-on	Ignore/Review
2)	Background	Defend/Lock-on	Ignore/Review
3)	Background	Defend/Lock-on	Ignore/Review
4)	Background	Defend/Lock-on	Ignore/Review
5)	Background	Defend/Lock-on	Ignore/Review
6)	Background	Defend/Lock-on	Ignore/Review
7)	Background	Defend/Lock-on	Ignore/Review
8)	Background	Defend/Lock-on	Ignore/Review
9)	Pre-Critical	Defend	Ignore
10)	Critical	Ignore	Defend
11)	Post-Critical	Defend	Ignore
12)	Background	Defend/Lock-on	Ignore/Review
13)	Background	Defend/Lock-on	Ignore/Review
14)	Background	Defend/Lock-on	Ignore/Review
15)	Background	Defend/Lock-on	Ignore/Review
16)	Background	Defend/Lock-on	Ignore/Review
17)	Background	Defend/Lock-on	Ignore/Review
18)	Background	Defend/Lock-on	Ignore/Review

[a]Session 1 nests the critical aircraft (low threat) in a series in which most aircraft were highly threatening.

[b]Session 2 nests the critical aircraft (high threat) in a series in which most aircraft were non-threatening.

all very similar to each other. In the first session, individuals and teams responded to 18 different signals in a hostile environment, where the first 9 incoming aircraft were very threatening (and hence required "defend" or "lock-on" judgments). The 10th signal (which represented a non-threatening aircraft) was the critical signal and it occurred roughly 35-40 minutes into the session. This timing was chosen based upon research indicating the traditional vigilance decrement manifested itself and stabilized within this time period [Teichner, 1974].

The second vigilance session (which occurred 3 weeks later) had a similar structure, however, it was a benign environment where the vast majority of aircraft were non-threatening. The 10th signal was very threatening, however, and this served as the critical signal.

There was no theoretical reason to expect that the nature of the environment (i.e., hostile or benign) would make a difference in the psychological processes involved and empirical checks showed no significant differences across environments in terms of decision making accuracy. Therefore, we collapsed the data across the nature of the environment (hostile or benign) yielding two observations for each team on the pre-critical, critical, and post-critical aircraft. Thus, the nature of the critical aircraft in terms of its threat level (threatening or non-threatening) is not confounded with its status as a pre-critical, critical or post-critical aircraft.

3. 2 Decision Accuracy

The dependent variable in this design is decision accuracy. Decision accuracy was obtained at the individual level for each of the three team staff members, and therefore, the sample size was 60 at the individual level (i.e., 3 staff members in each of 20 teams). It was also obtained for the team, where the sample size was 20.

4. Data Analysis and Statistical Power

Repeated measures regression was used to analyze the data [Cohen & Cohen, 1983]. [Hollenbeck, Ilgen and Sego, 1994] show how this technique can be applied to studies of teams and discuss its advantages in terms of statistical power. The technique first requires partitioning the overall variance in the dependent variable into within and between subjects variance, and then it systematically analyzes each portion separately. Statistical power is gained by obtaining multiple observations per team and by removing irrelevant sources of variance from the denominator of the F-ratio when making inference tests (e.g., removing between team variance when examining within team effects).

5. Results

5. 1 Hypotheses 1 and 2

At the individual staff member level, Hypotheses 1 and 2 are tested with a regression analysis where the dependent variable is the accuracy of the staff members' recommendations, and the independent variables are two dummy coded variables. The dummy coding scheme treats the pre-critical aircraft as the control group (i.e., coded 0,0). The first dummy variable contrasts decision accuracy on the critical aircraft (coded 1,0) with accuracy on the pre-critical control aircraft, and the second dummy variable contrasts the post-critical aircraft (0,1) with the pre-critical control aircraft. A statistically significant effect for the first dummy variable would support Hypothesis 1 by indicating a performance difference between the pre-critical aircraft and the critical aircraft. A statistically significant effect for the second dummy variable would support Hypothesis 2, by indicating a performance difference between the pre-critical aircraft and the post-critical aircraft.

These regressions are based on 180 observations, because each of 3 staff members in each of 20 teams is observed over 3 different conditions. This number of observations provides a power of .99 to detect a moderate effect size (i.e., $r = .35$) for either dummy variable.

The results of these regressions, in terms of variance partitioning, are depicted in the left hand side of Figure 1. At the individual level, 40% of the total variance in the staff members' accuracy was explained by the nature of the aircraft in terms of it being the pre-critical, critical or post-critical signal. Since 75% of the total variance was attributable to within-person variance (versus 25% attributable to between-person variance), this means that a statistically significant

53% of the within-person variance could be explained by the nature of the aircraft.

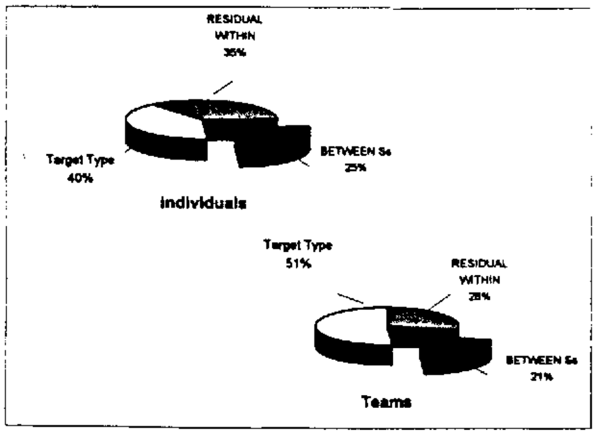

Figure 1. Within subject variance accounted for by type of signal in individuals and teams.

The nature of these effects is revealed in the left hand side of Figure 2, which shows the average level of decision accuracy for the three types of aircraft for the staff members. Since accuracy is defined as the difference between the "true decision" and the actual decision, lower scores reflect greater accuracy. The figure reveals that the staff members were less accurate on both the critical aircraft and the post-critical aircraft relative to the pre-critical control aircraft, supporting both Hypotheses 1 and 2.

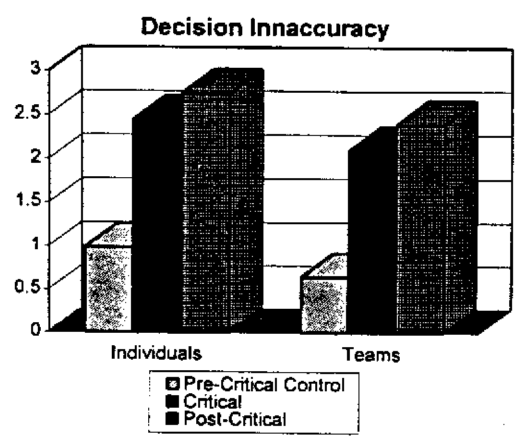

Figure 2. Average level of decision accuracy for pre-critical control, critical, and post-critical signals among individuals and teams.

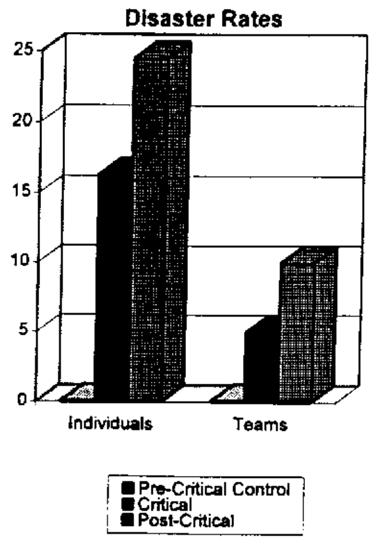

Figure 3. Disaster rates for pre-critical control, critical, and post-critical signals for individuals and teams.

Whereas Figure 2 depicts the results in terms of the continuous measure of decision accuracy, another way to depict the effects is to examine the number of highly inaccurate decisions (i.e., disasters) that occurred under different conditions.

The left hand side of Figure 3 shows that no staff member's recommendation regarding the pre-critical control aircraft would have lead to a disaster. However, 16% of their recommendations regarding the critical aircraft would have lead to a disaster, and 23% of the their recommendations would have led to a disaster on the post-critical aircraft.

5. 2 Hypothesis 3

The procedure to test Hypotheses 3 is identical to what was described above, but the number of observations is reduced. At the team level, there is only one decision per team, although each team is still observed under three different conditions. The total number of observations (i.e., 60) provides a power of .75 to detect a moderate effect size ($r = .35$).

The results from this analysis are shown in the right hand side of Figure 1. Fifty-one percent of the total variance in team decision accuracy is explained by the nature of the aircraft. Since 79% of the total variance was attributable to within team variance (versus 21% attributable to between team variation), this means that a statistically significant 65% of the within team variance is accounted for by the nature of the aircraft.

The nature of these effects at the team level is shown in the right hand side of Figure 2. Team decision accuracy was lower for both the critical and post-critical aircraft relative to the pre-critical control aircraft. The right hand side of Figure 3, also shows these results plotted in terms of severely inaccurate decisions (i.e., disasters). No teams experienced a disaster with the pre-critical control aircraft, but the disaster rate for the critical aircraft and the post-critical aircraft was 5% and 10%, respectively. Thus, the performance failures exhibited by individual staff members (tested in Hypotheses 1 and 2) seemed to carry over to teams as a whole, supporting our third hypothesis.

6. Discussion

This study examined human performance in contexts requiring sustained attention, and further documented the types of performance failures that often occur on tasks that require extensive passive monitoring. Given the increased role that passive monitoring plays in many tasks due to technological changes in the nature of work [Human Capital Initiative, 1993], this is a topic that should be of critical interest to all researchers in applied psychology.

The present study replicated the vigilance decrement typically found in human factors experiments with a new simulation task. One of the most important novel feature of this task was the ability of the research participants to respond on a graduated scale. Traditional tasks in vigilance experiments simply allow respondents a dichotomous choice. Use of the graduated scale, however, allowed us to see that, although the individuals in our study failed to respond appropriately to the critical signal, they rarely responded with a decision that would have been appropriate for the background signal. Instead, a large percentage took compromise positions like ready, warn, or monitor.

This suggests that many people perceived that there was "something different" about the incoming critical aircraft, and adjusted their responses. They did not adjust enough, however. Although most of the individuals and teams avoided outright disasters, most still wound up misclassifying the aircraft.

This inference was supported by an examination of response latency data also available from the simulation. These data indicated that the critical aircraft did not just slip by these observers undetected. The individual staff members took, on average, 7 seconds longer to process the critical aircraft relative to the pre-critical aircraft. On average, the leaders took 5 seconds longer. It appeared that the subjects could see it, they just did not seem to believe it (at least not completely). Thus, while information processing theory suggests both attentional and judgmental reasons for problems in sustained attention contexts, our data suggests that the latter may be more critical than the former.

Beyond replicating a vigilance-type decrement with a different task, this study also extended this literature by documenting a post-critical signal decrement. For the individuals involved in our study, the performance decrement associated with a return to the "normal" background aircraft after experiencing a non-routine event was even greater (in an absolute sense) than the original decrement associated with the critical signal. This problem appears even more extreme when one looks at disaster rates for individuals and teams. This result conforms with past research on the availability heuristic which shows that recent, unusual, and easy to recall events have a disproportionate effect on individuals' judgments and decisions. This kind of dynamic effect, where past outcomes reach forward to affect future responses is an under-researched area in the area of team decision making [Stevenson, et al. 1991].

Within organizations, any conspicuous error is likely to generate attention and remedial measures aimed at preventing a problem of the same type in the future. Whereas these remedial measures seem to make sense, in reality, they may add to an already existing pre-disposition to over-correct in these type of situations. This may significantly increase the probability of the opposite type of error.

For example, after the U.S.S. Stark incident (in which 37 servicemen died on a U.S. vessel that failed to defend itself against a threatening aircraft), the rules of engagement were tightened up to prevent the re-occurrence of this type of error. In the words of then-President Ronald Reagan, "From now on, if aircraft approach any of our ships in a way that appears hostile, there is one order of battle --

defend yourselves, defend American lives" [Jacoby, 1987, p. 17]. This sentiment was also expressed by then-Secretary of Defense Caspar Weinberger who stated that U.S. ships following that incident were operating "under hair trigger alert, prepared to fire on any plane that approaches in a hostile manner" [Lamar, 1987, p. 13]. Shortly after this statement was made, the U.S.S. Vincennes mistakenly shot down a passenger plane that was misjudged to present a threat. A similar, more recent incident occurred within the context of security around the White House in 1994. After two separate incidents of someone firing gunshots toward the White House, a host of agents from the U.S. Parks Service surrounded and fatally shot a homeless vagrant who was carrying a small knife near Pennsylvania Avenue.

Our research also extended the previous literature on decision making by examining the degree to which problems in sustained attention found among individuals decision makers would extend to situations where a decision was made by a team. One could speculate that teams might be less likely to miss the critical signal. The increased number of observers available to spot the incongruity between the critical signal and the background signal might reduce the probability of error.

Indeed, qualitative data that could be obtained from this simulation seemed to indicate that this effect could occur. That is, we isolated the team that performed best on the two critical targets and analyzed every text message sent between team members. With 59 seconds remaining in Trial 10 (the critical target) in the first vigilance session, one team member (i.e., CAD) announced, "Ready to Lock-on...How 'bout you?." This judgment would have lead to a disaster. However, with 51 seconds remaining in the trial, a second team member (i.e., Cruiser) correctly noted that "Its far away and not coming at us." These two factors, given the interactive nature of the rules, make the target non-threatening.

On the other hand, the data as a whole suggested that this type of effect was the exception rather than the rule. Instead, it appeared that social loafing or the distraction caused by the presence of other people prevented teams from successfully prosecuting the critical aircraft. For example, an examination of the text messages sent between members of the team that performed most poorly on the critical target, indicated that although half of their text messages were task-related on the first five trials, by the time the tenth trial came along, all the text messages were social as opposed to task-related. Thus, when confronted with a task that is perceived as boring, the presence of others provided a source of distraction that ultimately led to a performance breakdown on the critical signal.

Thus, taken as a whole, the results of this study provide no support for the notion that placing this sort of task within a team context eliminates problems with sustained attention. Although teams were slightly better overall than individuals on the pre-critical, critical and post-critical aircraft, because of the lower variability of team performance (the standard deviation of decision accuracy for teams was 1.02 compared to 1.23 for individuals), the overall effect in terms of variance accounted for by the type of aircraft was actually greater for teams relative to individuals (51% versus 40%). Since combining decision makers into teams fails to eliminate problems in the area of sustained attention, future research needs to explore other avenues for redressing the types of decision errors that occur in these contexts.

7. References

[Anderson, J.R., 1983]. The retrieval of information from long-term memory. *Science, 20*, 25-30.

[Campbell, J.P., 1991]. Modeling the performance prediction problem in industrial and organizational psychology. In M.D. Dunnette & L.M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology* (pp. 687-732). Palo Alto, CA: Consulting Psychologists Press.

[Cohen, J., & Cohen, P., 1983]. *Applied multiple regression/ correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

[Collins, B.E., & Guetzkow, H., 1964]. *A social psychology of group processes for decision making*. New York: Wiley.

[Fisk, A.D., Ackerman, P.L., & Schneider, W., 1987]. Automatic and controlled processing theory and its application to human factors problems. In P.A. Hancock

(Ed.), *Human factors psychology*. North-Holland: Elsevier Science Publishers.

[Hollenbeck, J.R., Ilgen, D.R., Phillips, J.M., & Hedlund, J., 1994]. Decision risk in dynamic contexts: Beyond the status quo. *Journal of Applied Psychology, 79*, 592-597.

[Hollenbeck, J.R., Ilgen, D.R., and Sego, D.J., 1994]. Repeated measures regression and mediational tests: Enhancing the power of leadership research. *Leadership Quarterly, 5*, 3-23.

[Hollenbeck, J.R., Ilgen, D.R., Sego, D.J., Hedlund, J., Major, D.A., and Phillips, J.M. (in press)]. The multi-level theory of team decision-making: Decision performance in teams incorporating distributed expertise. *Journal of Applied Psychology*.

[Hollenbeck, J.R., Sego, D.J., Ilgen, D.R. & Major, D.A., 1991]. Team interactive decision exercise for teams incorporating distributed expertise: A program and paradigm for team research. Technical Report 91-1, East Lansing, MI: Michigan State University.

[Hollenbeck, J.R., Sego, D.J., Ilgen, D.R., Major, D.A., Hedlund, J., and Phillips, J.M. (in press)]. Team decision making accuracy under difficult conditions: Construct validation of potential manipulations and measures using the $TIDE^2$ simulation. In M. Brannick and E. Salas (Eds.), *New Directions in Team Measurement*. San Francisco: Jossey Bass.

[Human Capital Initiative, 1993]. Making people and technology work well together. Human Capital Initiative: The changing nature of work, *American Psychological Society Observer*, October 1993, 1-65.

[Janis, I.L., 1982]. *Groupthink*, Boston, MA: Houghton-Mifflen.

[Jacoby, T., 1987]. A tragedy in the gulf. *Newsweek*, June 1, 1987, 16-19.

[Kessel, C.J., & Wickens, C.D., 1982]. The transfer of failure-detection skills between monitoring and controlling dynamic systems. *Human Factors, 24*, 49-60.

[Lamar, J., 1987]. Rough seas and new names. *Time*, June 29, 1987, p. 13.

[Latane, B., Williams, K., & Harkins, S., 1979]. Many hands make light work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology, 37*, 822-832.

[Laughlin, P.R., 1980]. Social combination processes of cooperative problem solving groups on verbal intellective tasks. In M. Fishbein (Ed.), *Progress in social psychology*, (pp. 127-156). Hillsdale, NJ: Erlbaum.

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

[Michaelsen, L.K., Watson, W.E., Black, R.H., 1989]. A realistic test of individual versus group consensus decision making. *Journal of Applied Psychology, 74*, 834-839.

[National Transportation Safety Board, 1984]. Scandinavian Airline Systems DC-10-30, John F. Kennedy International Airport, Jamaica, New York, February 28, 1984 (Report NTSB-AAR-84-15). Washington D.C.: NTSB.

[National Transportation Safety Board, 1986]. China Airlines Boeing 747-SP, N4522V, 300 nautical miles northwest of San Francisco, CA, February 19, 1986 (Report NTSB-AAR-86-03). Washington D.C.: NTSB.

[Schneider, W., & Shiffren, R.M., 1977]. Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review, 84*, 1-66.

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

[Stevenson, M.K., Busemeyer, J.R., & Naylor, J.R., 1991]. Judgment and decision making theory. In M.D. Dunnette & L.M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology*. Palo Alto, CA: Consulting Psychologists Press.

[Teichner, W.H., 1974]. The detection of a simple visual signal as a function of time of watch. *Human Factors, 16*, 339-353.

[Tinsdale, R.S. & Larson, J.R., 1992]. Assembly bonus effect or typical group performance? A comment on Michaelsen, Watson, and Black (1989). *Journal of Applied Psychology, 77*, 102-105.

[Tversky, A. & Kahneman, D., 1981]. The framing of decisions and the psychology of choice. *Science, 211*, 453-458.

[Wiener, E.L., 1987]. Application of vigilance research: Rare, medium, or well done? *Human Factors, 29*, 725-736.