April 1992

GMU/C3I-220-R

COGNITIVE BIASES AND TIME STRESS IN TEAM DECISION MAKING

Paul Lehner, Mir-Masood Seyed-Solorforough,
Bhashyam Nallappa, Michael O'Conner,
Stephen Sak and Theresa Mullin

Center of Excellence in Command,
Control, Communications, and Intelligence (C^3I)

George Mason University
Fairfax, Virginia 22030

April 1992 GMU/C3I-220-R

COGNITIVE BIASES AND TIME STRESS IN TEAM DECISION MAKING

Paul Lehner, Mir-Masood Seyed-Solorforough, Bhashyam Nallappa, Michael O'Conner, Stephen Sak and Theresa Mullin

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704 0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
April 1992

3. REPORT TYPE AND DATES COVERED
Technical Report, 12/91 to 3/92

4. TITLE AND SUBTITLE

COGNITIVE BIASES AND TIME STRESS IN TEAM DECISION MAKING

5. FUNDING NUMBERS

N00014-90-J-1680

6. AUTHOR(S)
Paul E. Lehner, Mir-Masood Seyed-Solorforough, Bhashyam Nallappa, Michael O'Connor, Steve Sak and Theresa Mullin

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Center of Excellence in Command, Control, Communications and Intelligence
George Mason University - Fairfax, Virginia 22030

8. PERFORMING ORGANIZATION REPORT NUMBER

GMU/C3I-220-R

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
Arlington, Virginia 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This experiment investigates the impact of time stress on the decision making performance of command and control teams. Two person teams were trained in a set of simple decision procedures. Some of these procedures required subjects to make judgments that were counter to normal heuristic decision processing. The principle hypothesis was that these decision procedures would be vulnerable-to-bias, and would therefore be more vulnerable to the effects of time stress than other decision procedures. The results support this hypothesis. In addition, the results suggest that the subjects adapted inappropriately to time stress. As time stress increased, they began to use a decision processing strategy that was less effective than the strategy they were trained to use.

14. SUBJECT TERMS

Team Decision Making - Cognitive Bias

15. NUMBER OF PAGES
21

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

COGNITIVE BIASES AND TIME STRESS IN TEAM DECISION MAKING

Paul Lehner, Mir-Masood Seyed-Solorforough and Bhashyam Nallappa

Center of Excellence in C3I
George Mason University
4400 University Drive
Fairfax, VA 22030

Michael O'Conner, Stephen Sak and Theresa Mullin

Decision Science Consortium
1840 Preston White Drive
Reston, Virginia 22071

ABSTRACT[1]

This experiment investigates the impact of time stress on the decision making performance of command and control teams. Two person teams were trained in a set of simple decision procedures. Some of these procedures required subjects to make judgments that were counter to normal heuristic decision processing. The principal hypothesis was that these decision procedures would be *vulnerable-to-bias*, and would therefore be more vulnerable to the effects of time stress than other decision procedures. The results support this hypothesis. In addition, the results suggest that the subjects adapted inappropriately to time stress. As time stress increased, they began to use a decision processing strategy that was less effective than the strategy they were trained to use.

[1]This research was funded by the Office of Naval Research (#N00014-90-J-1680) and the Center of Excellence in C3I at George Mason University. The Center's general research program is sponsored by the Virginia Center for Innovative Technology, the MITRE Corporation, the Defense Communications Agency, CECOM, PRC/ATI, ASD(C3I), TRW, AFCEA, and AFCEA NOVA.

1.0 INTRODUCTION

This experiment is part of a research effort that is leading toward a prescriptive theory for deriving team decision making procedures (Lehner, 1991). In this prescriptive theory a set of decision procedures are derived from a decision theoretic domain model of the decision situations that a decision making team is expected to face. The approach to deriving a team member's decision procedures attempts to achieve three objectives: high expected performance, reasonable workload, and consistency with normal human decision making procedures. This last objective is based on the hypothesis that "unnatural" decision procedures are not likely to be reliably executed under conditions of time stress. The experiment reported here is a test of this hypothesis.

1.1 Team Decision Making

A *team* is defined as a group of decision makers, with overlapping areas of expertise, that work cooperatively to solve common decision problems. The cockpit crew of a commercial aircraft (pilot, co-pilot, navigator) is an example of small team, while the crew of a submarine control room is an example of a somewhat larger team. Command and control teams (C^2 team) are further characterized as teams where (a) each member of the team is responsible for an assigned set of tasks, although the assignment of tasks may change dynamically, (b) each team member has been trained for the tasks that he or she is to perform, and (c) for the problems the team faces, the team members have the common goal of satisfactorily solving that problem. Finally, we note that C^2 teams often address problems that are severely time constrained.

C^2 teams are generally well-trained. Consequently, C^2 team performance is usually high. However, "human error" does sometimes occur. This is particularly true under conditions of stress. If the consequences of decisions are severe, and there is very little time for the team to perform its tasks, catastrophic mistakes can easily occur. The objective of this research is to experimentally identify team decision procedures that tend to be executed relatively unreliably under conditions of time stress.

1.2 Cognitive Biases and Behavioral Decision Theory

Behavioral Decision Theory (BDT) is a branch of psychology that performs research in human judgment and decision making (JDM). Although related to cognitive psychology research in such areas as human problem solving (e.g., Newell & Simon, 1972), memory, natural language understanding, and the like, it has traditionally been separated from these areas by two features (Lehner and Adelman, 1990). First, there is an emphasis on tasks that involve quantitative tradeoffs and the integration of subjective judgments (e.g., probability assessment, option selection). Second, there is more of a focus on models that characterize JDM performance rather than JDM processes. For instance, BDT researchers often use linear models to characterize behavior in contexts where it is believed that the "internal" JDM process is pattern-based and non-linear.

An important result in the BDT literature is that human decision making behavior exhibits a number of *cognitive biases*. Cognitive biases are defined as judgments that consistently deviate from a normative ideal. The standard explanation for the occurrence of cognitive biases is that people employ a variety of heuristic procedures when making judgments that bear little resemblance to theoretically normative procedures. Some of these biases are listed below.

> *Availability Bias*. People often overestimate the probability of an event that is easy to recall or imagine (e.g., Tversky & Kahnemann, 1973).
>
> *Confirmation Bias*. People tend to seek and focus on confirming evidence, with the result that once they've formed a judgment, they tend to ignore or devalue disconfirming evidence (e.g., Wason, 1960; Tolcott, et.al., 1989).
>
> *Frequency Bias*. People often judge the strength of predictive relations by focusing on the absolute frequency of events rather than their observed relative frequency. As Einhorn & Hogarth (1978) have shown, information on the nonoccurrence of an event is often unavailable and frequently ignored when available.
>
> *Concrete Information*. Information that is vivid or based on experience or incidents dominates abstract information, such as summaries or statistical base-rates. According to Nisbett & Ross (1980), concrete and vivid information contributes to the imaginability of the information and, in turn, enhances its impact on inference.
>
> *Conservatism*. If people are forced to consider base rates, then they often underestimate the predictive value of new information. That is, their revised probability estimates remain too close to the original base rates (Edwards, 1968).

Anchoring and Adjustment. A common strategy for making judgments is to anchor on a specific cue or value and then to adjust that value to account for other elements of the circumstance. Usually the adjustment is insufficient. So once the anchor is set, there is a bias toward that value (Kahneman & Tversky, 1973).

"Law of Small Numbers". Problems can be framed in such a way that people, including trained statisticians, give undue confidence to conclusions supported by a relatively small amount of data (Tversky & Kahneman, 1971).

Hindsight Bias. This is perhaps the most problematic of all biases (Fischhoff, 1975). After an event occurs, people will often claim they predicted the event ("I knew it all along!"), even though prior to the event they were very uncertain.

Fundamental Attribution Error. People tend to attribute success to their own skill and failure to chance or the circumstances in which they were situated. However, when evaluating the performance of other people, the tendency is to attribute other people's failure to their personality traits, not the situation (see Nisbett & Ross, 1980).

Because of the prevalence of cognitive biases, there has also emerged a research literature addressing the "debiasing" problem. The objective of this research is to develop presentation and problem solving techniques that reduce biases. This literature is reviewed in O'Conner (1992). Despite a few successes, the principal result seems to be cognitive biases are very resistant to debiasing techniques.

1.3 Cognitive Biases and Team Decision Making

The principal focus of the cognitive bias literature is on natural decision making procedures. Consequently, it can be argued that cognitive biases are irrelevant to C^2 team decision making, since team members are well-trained in their decision making tasks. This argument asserts that, as long as workload is not excessive, people will reliably execute whatever decision procedures they have been trained to execute; irrespective of what they might naturally do if they weren't trained.

In contrast, one could argue that "unnatural" decision procedures are *vulnerable-to-bias*. A vulnerable-to-bias procedure is defined as a judgment or decision procedure that an untrained or poorly trained decision maker would not reliably execute, because of a cognitive bias. For instance, a judgment procedure that requires a person to make an initial judgment based on sparse evidence, and then to update that judgment based on additional evidence, is vulnerable to both the conservatism and the confirmation bias, as well as to any other biases that might result from the use of an anchoring and adjustment heuristic. These biases suggest that people would normally undervalue the impact of new evidence, and that they may stick with their original judgment in the face to strong counter evidence.

Vulnerable-to-bias procedures may impact team decision making in two ways. First, it may be difficult to train decision makers in decision procedures that are vulnerable-to-bias. The debiasing literature (O'Conner, 1991) suggests that this is likely to occur. Second, under conditions of stress, it is possible that vulnerable-to-bias decision procedures will not be as reliably executed as more natural decision procedures. It is this later possibility that is investigated here.

The experiment reported below contrasts two perspectives.

> *Perspective 1 (P1)* - Cognitive biases are largely a matter of preference. Although people tend to use heuristic rules that deviate from normative decision theory, they can be taught to reliably use alternative rules; as long as the alternative rules do not exceed workload constraints.

> *Perspective 2 (P2)* - Cognitive biases are largely a matter of capability. Even after training, people do not reliably execute judgment and decision procedures that are vulnerable-to-bias.

For team decision making under stress, these two perspectives differ considerably with respect to their implications for team design. If P1 is correct, then the literature on human cognitive biases is largely irrelevant to the problem of designing teams. Properly trained and practiced teams will reliably execute correct decision procedures until workload or other bounded rationality constraints are exceeded. On the other hand, if P2 is the correct perspective, then cognitive bias considerations should place severe constraints on the design of a team. Specifically, team architectures and decision procedures should avoid vulnerable-to-bias decision procedures, since these decision procedures will not be reliably executed in high stress conditions.

2.0 METHOD

The experiment was a modification of that of Jin (1990). Two person teams worked together to defend a battle group from incoming aerial attacks. Each team member must assess the type of aircraft associated with each radar track.

2.1 Subjects

Subjects were paid volunteers from the graduate and undergraduate student population at George Mason University. Data from eleven two person teams was collected. In addition, pilot data from four two person teams was collected.

2.2 Materials

This game was played with two players and three Macintosh LC computers. Each player was placed behind one computer. The players were named DM1 and DM2 on the computer. The third computer was named DM3. In this experiment the DM3 computer was only used to synchronize the game between between DM1 and DM2.

Display.

The display was divided into six windows. Figure 2-1 shows the display from DM1's perspective. The display and all procedures for DM2 is a mirror image of DM1.

The first window (upper left) was for displaying Threat Information. When a threat was selected, (see below) information about that threat was displayed. This information includes an ID number, the angle of attack on the screen (R), the velocity of the track (V), and the number of aircraft in the track (Num). In addition, if the other DM sent advice regarding the type of threat (see below) that information was also displayed (Type). However, this Type advice was not displayed until after the DM made an initial judgment as to the type of the aircraft.

The second window (below the first window) displays the resources assigned to each threat after the threat and the Attack button were selected. This region includes two buttons, Attack and Clear. The Clear button was not used in this game. The Attack button was routinely selected after target Type was selected.

The third window, which is located below the second region, is the Messages window. Messages from the other player were displayed in this region. In this experiment, the only information that was displayed was the other DMs assessment as to the type of aircraft in a track. This Type "advice" was only displayed after an initial Type judgment was made by DM1.

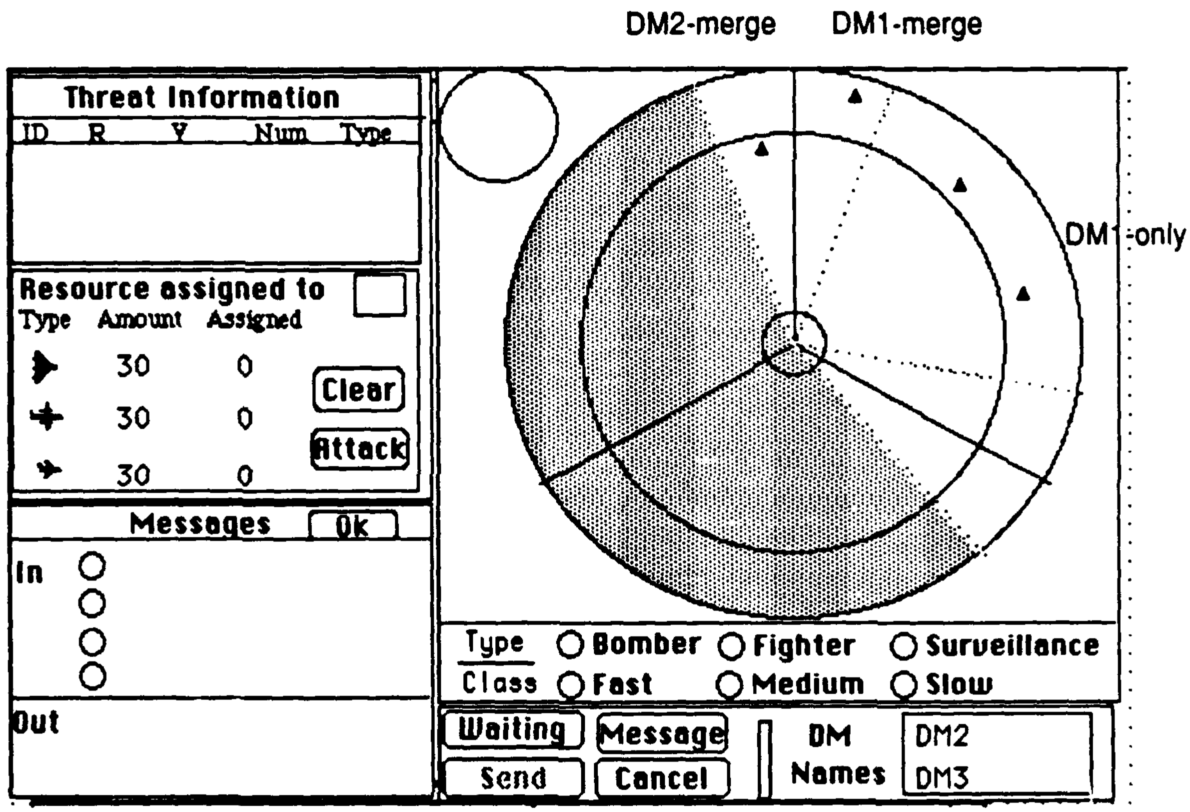

FIGURE 2-1: SCREEN DISPLAY FOR DM1

The fourth window is the Radar. The Radar display was located on the top right hand side of the screen. In this region the incoming threats were displayed by a small triangle and the players selected individual threats by placing the cursor on the threat and clicking the mouse button. Once this action is carried out, information about the threat was displayed in the Threat Information window. From DM1's perspective radar, the radar was partitioned into four regions. They are DM1-merge, DM2-merge and DM1-only and the region where no information was displayed. For the DM1-only region, DM1 could click on a track; make judgments with respect to its Class and Type; and hit the Attack button. For the DM2-merge region, DM1 could click on a track; make judgments with respect to its Class and Type; and then send that advice to DM2 by pressing the Send button. For the DM1-merge region, DM1 could click on a track; make judgments with respect to its Class and Type; review the advice from the other DM; change his or her Type judgment; and click the Attack button.

In the remainder of this report, the regions on the Radar screen will be generically referred to as DMi-only, DMi-merge, DMj-merge.

The fifth window is located below the radar screen. It is used to record Class (Slow, Medium, Fast), and Type (Bomber, Fighter, Surveillance) judgment for the threats that appear on the Radar. These are selected based on the threat information that appear in the Threat Information window and advice that appears in the Messages and Threat Information windows.

The sixth window is located at the bottom right hand side. In this window there are four keys: Wait, Message, Send and Cancel. This region was used to transmit type advice from one DM to the other. This was carried out by selecting the destination from the list of DMs, selecting the Message button and then selecting the Send button.

2.3 Procedures

Objective and Rules of the game

The team objective was to correctly identify and attack all the threats before they reached the small center circle. For all three regions, each DM must initially to use number and velocity information to select Class, Size and Type. The rules for these initial judgments were as follows.

Class

Slow	if Velocity < 500
Fast	if Velocity > 800
Medium	otherwise

Size

Small	if Number < 5
Medium	if Number = 5
Large	if Number >5

Type		
	Fighter	if Speed = Slow and Class = Small
		if Speed = Slow and Class = Medium
		if Speed = Fast and Class = Large
	Bomber	if Speed = Slow and Class = Large
		if Speed = Medium and Class = Large
	Surveillance	otherwise.

For threats in the DMi-only or DMj-merge regions either the Attack or Send button was hit after the Type judgment was made. For the DMi-merge region, after Type was selected, DM2's advice was displayed, if any had been sent. At this point DMi was trained to revise his or her Type judgment according to the following rules.

Condition	Initial Type Judgment	Type Advice Received	Final Type Judgment
FF	Fighter	Fighter	Fighter
FB	Fighter	Bomber	Bomber
FS	Fighter	Surveillance	Surveillance
BF	Bomber	Fighter	Bomber
BB	Bomber	Bomber	Bomber
BS	Bomber	Surveillance	Fighter
SF	Surveillance	Fighter	Surveillance
SB	Surveillance	Bomber	Fighter
SS	Surveillance	Surveillance	Surveillance

Note that the rules for the FB and FS condition require the DM to modify his or her original opinion to conform with the new information (the other DM's advice). Also the BS and SB conditions require that the DM select a new Type that that was inconsistent with both DMs' initial selection. The decision procedure for the FB,FS and the BS,SB conditions were considered to be vulnerable-to-bias, since they required subjects to anchor and then adjust their original judgment.

Instructions.

Subjects were instructed about the general nature of the game. They were told that the tactical information for threats arriving in the merge (i.e., DM1-merge and DM2-merge) regions was not as reliable as the information in the DMi-only regions. Consequently, the rules for combining initial Type judgment with the other DM's advice were needed to increase accuracy.

Training.

The subjects were trained for a period of approximately 1.5 hours before the game started. Training occurred in four stages.

During the first stage, the players were trained in their DMi-only region. They were instructed to ignore tracks arriving in the other regions. This continued until the subjects were comfortable with the rule set. The training set involved tracks at varying speeds (see below).

Once the players were comfortable dealing with the threats in DMi-only region, they were instructed to ignore this region and concentrate only on the threats in the DMj-merge region, where they had to make a Type judgment and send that judgment as Type advice to the other DM. Once again, they repeated this (at varying speeds) until they were comfortable with the rules and procedure.

Third, the players were instructed to practice sending advice and then to deal with their DMi-merge region. Here they were required to learn the rules for adjusting their initial judgments on the basis of the other DM's advice. Once again this continued until the subjects were comfortable with the rule set and general procedure.

Finally, they were told to practice the complete game and try to stop all threats before they came to the center of the radar. It was recommended that subjects process tracks in the DMj-merge, DMi-only, and the DMi-merge regions respectively. This guaranteed that the other DM had time to send advice before the subject began processing his or her DMi-merge tracks.

This practice session was repeated until the players felt proficient with the rules.

In all of these practice sessions the players were allowed to look at two tables that summarized the judgment rules. This was done in order to let them confirm that their decisions were correct.

Experimental Tasks.

Each team processed 12 sets of ten trials each. The 12 sets were partitioned into four sets each with tracks moving at 400, 600, 750 and 900 mph respectively. At a speed of 400 mph the subjects had 45 seconds to process all tracks. At 600 mph, they had 30 seconds. At 750 and 900 mph, they h. d 24 and 20 seconds respectively. There was a short break between trials (approximately 20 seconds) and a longer break between sets of trials (about four minutes).

The training and experimental sessions were distributed over two days. Each team took a total of about 2.5 hours to complete all 12 sets of trials. Consequently, teams spent an average of about 4 hours in training and experimental sessions.

It should be noted that the Appletalk network crashed about once every 6 hours during training and experimentation. Consequently, it was sometimes necessary to have a short break during a set of trials to restart on the next track.

Each trail was composed of 8 tracks. Two in each of the four regions (DM1-only, DM1-merge, DM2-merge, DM2-only). Subjects were simply instructed to do their best to process as many of the incoming tracks as possible.

3.0 RESULTS

3.1 Reliability

An initial examination of the data revealed that for two of the teams, performance on the slowest speed trials was less than 50%. Since a performance level of 33% can be obtained by answering randomly, the data from these teams was not included in the analysis. The performance of the other nine teams, for the slowest speed, was consistently high.

3.2 Overall Accuracy

Tables 3-1 through 3-4 summarize the proportion of tracks that were correctly processed for each of the trial speeds. The columns on the four tables are defined as follows.

DMi-only: This column represents proportion correct in the DMi-only region.

Advice sent: This column represents the proportion of tracks in the DMj-merge region for which the correct advice was sent to the other decision maker.

No merge: For the tracks in the DMi-merge region for which advice was not received, this column represents the proportion of tracks for which the correct judgment was made based only on just the information in the Threat Information window.

Merged: For the tracks in the DMi-merge region for which advice was received, this column represents the proportion of times the correct final judgment was made.

Table 3-5 summarizes the average results for the four speed conditions. The performance in the No-merge group is largely irrelevant, since it reflects a condition were subjects did not receive an input they were expecting. Performance under this condition is expected to be low.

TABLE 3-1: PERFORMANCE BY REGION, SPEED = 400.

Subject	DMi-only	Advice sent	No merge	Merged
TM1DM1	0.97	0.98	-	0.98
TM1DM2	0.97	0.98	-	0.68
TM2DM1	1	1	0	0.97
TM2DM2	0.97	0.98	-	0.97
TM4DM1	1	0.98	1	0.98
TM4DM2	0.95	0.97	0.5	0.94
TM5DM1	0.95	1	0.38	0.88
TM5DM2	0.7	0.77	0.22	0.75
TM6DM1	0.97	1	-	0.95
TM6DM2	1	1	1	0.98
TM7DM1	0.97	1	0.75	0.93
TM7DM2	0.95	0.92	0.1	0.86
TM8DM1	0.84	0.57	0.61	0.71
TM8DM2	0.78	0.57	0.3	0.57
TM9DM1	1	0.98	0.29	0.94
TM9DM2	0.98	1	1	0.98
TM10DM1	1	1	1	1
TM10DM2	0.98	1	0.8	0.93
Overall Result	0.943	0.929	0.446	0.916

TABLE 3-2: PERFORMANCE BY REGION, SPEED = 600.

Subject	DMI-only	Advice sent	No merge	Merged
TM1DM1	0.86	0.95	0	0.96
TM1DM2	0.83	1	0.67	0.79
TM2DM1	1	0.82	0.65	0.84
TM2DM2	1	1	0.5	0.93
TM4DM1	1	0.97	0.6	0.98
TM4DM2	0.83	0.93	0.17	0.83
TM5DM1	0.88	0.97	0.34	0.92
TM5DM2	0.67	0.78	0.18	0.81
TM6DM1	0.82	1	-	0.9
TM6DM2	0.85	1	-	1
TM7DM1	1	1	0.69	0.94
TM7DM2	0.73	0.87	0.14	0.75
TM8DM1	1	0.97	0.74	1
TM8DM2	0.62	0.33	0.2	0.75
TM9DM1	0.97	0.88	0.32	0.71
TM9DM2	1	1	0.5	0.89
TM10DM1	0.98	1	0.17	0.96
TM10DM2	1	0.98	0.42	0.98
Overall Result	0.89	0.914	0.379	0.909

TABLE 3-3: PERFORMANCE BY REGION, SPEED = 750.

Subject	DMI-only	Advice sent	No merge	Merged
TM1DM1	0.82	1	0	0.96
TM1DM2	0.73	1	1	0.81
TM2DM1	0.95	0.98	0.28	0.79
TM2DM2	0.98	0.98	0.33	0.98
TM4DM1	0.98	0.98	0.46	0.87
TM4DM2	0.86	0.93	0.38	0.83
TM5DM1	0.52	0.67	0.26	0
TM5DM2	0.53	0.78	0.29	0.8
TM6DM1	0.54	0.98	1	0.88
TM6DM2	0.32	1	0.67	0.98
TM7DM1	0.98	1	0.54	0.81
TM7DM2	0.72	0.9	0.29	0.73
TM8DM1	1	0.97	0.49	0.67
TM8DM2	0.78	0.23	0.32	0
TM9DM1	1	0.81	0.23	0.85
TM9DM2	1	0.9	0.26	0.82
TM10DM1	0.95	0.98	0.22	0.93
TM10DM2	0.88	1	0.25	0.91
Overall Result	0.808	0.907	0.327	0.881

TABLE 3-4: PERFORMANCE BY REGION, SPEED = 900.

Subject	DMI-only	Advice sent	No merge	Merged
TM1DM1	0.48	1	0.26	0.94
TM1DM2	0.5	1	0.56	0.76
TM2DM1	0.77	0.98	0.33	0.67
TM2DM2	0.98	1	0.41	0.85
TM4DM1	0.9	1	0.28	0.6
TM4DM2	0.74	0.98	0.29	0.56
TM5DM1	0.37	0.73	0.29	
TM5DM2	0.47	0.78	0.26	0.33
TM6DM1	0.18	0.98	0.83	0.7
TM6DM2	0.08	1	0.43	0.98
TM7DM1	1	1	0.41	0.81
TM7DM2	0.6	0.9	0.32	0.8
TM8DM1	1	0.89	0.18	0
TM8DM2	0.75	0.42	0.25	1
TM9DM1	0.98	0.75	0.15	0
TM9DM2	0.98	0.72	0.32	0.33
TM10DM1	1	0.92	0.27	0.67
TM10DM2	0.93	1	0.3	0.7
Overall Result	0.69	0.889	0.295	0.766

TABLE 3-5: ACCURACY AT DIFFERENT SPEEDS

Speed	DMI-only	Advice sent	No merge	Merged
400	0.943	0.929	0.446	0.916
600	0.89	0.914	0.379	0.909
750	0.808	0.907	0.327	0.881
900	0.69	0.889	0.295	0.766

TABLE 3-6: AVERAGE RANK ORDER OF PROCESSING OF TRACKS

	DMI-only		DMj-merge		DMI-merge	
Speed	Mean	S.D.	Mean	S.D.	Mean	S.D.
400	3.32	1.36	2.07	0.84	4.55	1.45
600	3.19	1.25	2.03	0.80	4.46	1.29
750	3.14	1.18	1.85	0.88	4.47	1.02
900	2.74	1.17	1.89	0.74	2.25	1.16

An examination of the other columns reveals that subjects consistently maintained high performance on the tracks where they needed to send advice. They also managed to maintain relatively high performance for the Merged judgments. Performance for the DMi-only judgments dropped more substantially as time available decreased. An detailed examination of the subject output files reveals that this is due to the order in which subjects processed the tracks. Under the lower time stress conditions, subjects would process tracks in the order recommended (i.e., send advice first, DMi-only tracks second, and then do the merge judgments). However, as time stress increased, subjects shifted the order. They would process the merge tracks before attempting to process the DMi-only tracks. This is reflected in Table 3-6, which indicates the average rank order of processing of the tracks in different regions.

3.3 Vulnerable-to-bias judgments

The principal hypothesis was that vulnerable-to-bias decision procedures are executed less reliably than other decision procedures and are more vulnerable to the effects of time stress. This implies that performance for the FF,BB,SS and the BF,SF conditions will be higher than for the FB,FS and the BS, SB conditions. Tables 3-7 through 3-10 summarize performance under each condition for each level of time stress. In general, performance at each level supported the hypothesis that the vulnerable-to-bias decision procedures were executed less reliably than the other decision procedures.. A statistical analysis of these results is reported in Table 3-11. This analysis was performed by examining the proportion of individual subjects for whom the relative performance results were in the predicted direction. For instance, for the prediction BF,SF > FB,FS at Speed=400, there were nine subjects for which the average performance for the BF and SF condition was higher than for the FB and FS conditions, seven with equal performance, and one with performance in the reverse. A one-tailed sign test on a ratio of 9:1 gives the result $p < .05$.

It should be noted, however, that the results do not directly support the notion that the vulnerable-to-bias procedures degrade more rapidly than other procedures as stress increased. Although the FF,BB,SS and BF,SF conditions were more reliably executed than the FB,FS and the BS,SB conditions, the difference in performance between these conditions did not increase consistently as time stress increased. However, this lack of support is an artifact of the experimental design, since it must be the case that in non time stressed situations subjects could flawlessly execute the simple decision procedures in this experiment.

3.4 Comparison to Pilot Study

A pilot study of four teams was run prior to executing the main experiment. The pilot procedure differed from the main experiment in that there was more training, subjects went through more sessions, and the slowest speed was considerably slower. In the pilot study, all four groups attained near perfect performance at the slowest speeds. The other results were about the same.

TABLE 3-7: PERFORMANCE OF DIFFERENT TYPES OF MERGE JUDGMENTS, SPEED = 400.

Subject	Not Vulnerable-to-bias FF,BB,SS	BF,SF	Vulnerable-to-bias FB,FS	BS,SB
TM1DM1	1	1	1	0.944
TM1DM2	1	0.545	0.417	0.611
TM2DM1	1	1	0.917	0.944
TM2DM2	1	1	1	0.889
TM4DM1	1	1	1	0.933
TM4DM2	1	1	1	0.833
TM5DM1	0.941	1	0.846	0.769
TM5DM2	0.733	0.7	1	0.625
TM6DM1	1	1	0.917	0.889
TM6DM2	1	1	1	0.944
TM7DM1	1	0.917	0.769	1
TM7DM2	0.941	1	0.636	0.857
TM8DM1	1		1	0
TM8DM2	1	0.667	0.5	0.25
TM9DM1	1	1	0.833	0.938
TM9DM2	0.941	1	1	1
TM10DM1	1	1	1	1
TM10DM2	1	1	0.917	0.813
Overall Result	0.9751773	0.93854749	0.88359788	0.86131387

TABLE 3-8: PERFORMANCE OF DIFFERENT TYPES OF MERGE JUDGMENTS, SPEED = 600.

Subject	Not Vulnerable-to-bias FF,BB,SS	BF,SF	Vulnerable-to-bias FB,FS	BS,SB
TM1DM1	1	1	1	0.867
TM1DM2	1	0.786	0.643	0.765
TM2DM1	1	1	0.4	1
TM2DM2	1	1	0.786	0.938
TM4DM1	1	1	0.889	1
TM4DM2	0.875	0.714	1	0.75
TM5DM1	1	1	1	0.667
TM5DM2	1	1	0.6	0.75
TM6DM1	1	0.75	0.917	0.867
TM6DM2	1	1	1	1
TM7DM1	1	1	0.8	0.9
TM7DM2	0.571	0.75	0.857	0.833
TM8DM1	1	1	1	-
TM8DM2	0	1	1	1
TM9DM1	1	1	0	0.667
TM9DM2	1	0.875	0.778	0.923
TM10DM1	1	1	0.9	0.929
TM10DM2	1	1	1	0.923
Overall Result	0.97252747	0.93006993	0.84	0.88505747

TABLE 3-9: PERFORMANCE OF DIFFERENT TYPES OF MERGE JUDGMENTS, SPEED = 750.

Subject	Not Vulnerable-to-bias FF,BB,SS	BF,SF	Vulnerable-to-bias FB,FS	BS,SB
TM1DM1	1	1	0.875	0.938
TM1DM2	1	0.667	0.583	0.882
TM2DM1	1	1	0.5	0.667
TM2DM2	1	1	0.875	1
TM4DM1	1	1	0.75	0.8
TM4DM2	1	1	-	0.667
TM5DM1	-	-	-	0
TM5DM2	0.667	-	-	1
TM6DM1	0.941	1	0.917	0.722
TM6DM2	1	1	1	0.941
TM7DM1	1	1	0.571	0.733
TM7DM2	1	0.667	0	1
TM8DM1	0.5	1	-	-
TM8DM2	-	-	0	-
TM9DM1	1	1	0.5	0.8
TM9DM2	1	1	0.667	0.667
TM10DM1	1	1	0.667	0.889
TM10DM2	1	1	0.857	0.778
Overall Result	0.97916667	0.94565217	0.74444444	0.82894737

TABLE 3-10: PERFORMANCE OF DIFFERENT TYPES OF MERGE JUDGMENTS, SPEED = 900.

Subject	Not Vulnerable-to-bias FF,BB,SS	Not Vulnerable-to-bias BF,SF	Vulnerable-to-bias FB,FS	Vulnerable-to-bias BS,SB
TM1DM1	1	0.889	1	0.875
TM1DM2	1	0.667	0.571	0.733
TM2DM1	1	0.5	-	-
TM2DM2	1	1	1	0.667
TM4DM1	0.75	0.667	0.667	0
TM4DM2	1	1	0.25	0.5
TM5DM1	-	-	-	-
TM5DM2	1	0.25	-	0
TM6DM1	0.933	0.833	0.545	0.5
TM6DM2	1	1	0.9	1
TM7DM1	0.8	1	0	0.778
TM7DM2	-	0	1	1
TM8DM1	-	0	-	0
TM8DM2	-	1	-	1
TM9DM1	-	0	0	0
TM9DM2	-	1	0	0
TM10DM1	0.5	1	0	0.667
TM10DM2	1	0.75	1	0.333
Overall Result	0.94285714	0.7625	0.67307692	0.68539326

TABLE 3-11: NUMBER OF SUBJECTS FOR WHICH VULNERABLE-TO-BIAS PROCEDURES HAD MORE ERRORS THAN THE NON VULNERABLE-TO-BIAS PROCEDURES.

	Predicted Direction of Effect											
	FF,BB,SS=BF,SF			BF,SF>FB,FS			BF,SF>BS,SB			FB,FS=BS,SB		
	Observed			Observed			Observed			Observed		
Speed	>	=	<	>	=	<	>	=	<	>	=	<
400	4	10	3	9	7	1*	13	2	2**	11	2	5
600	5	11	2	10	5	3*	9	4	4	5	2	9
750	2	11	2	12	1	0***	11	1	2*	3	1	9
900	7	3	2	7	3	3	10	4	2*	5	3	5

* $p < .05$ (one-tailed sign test)
** $p < .01$
*** $p < .001$.

4.0 DISCUSSION

Overall the results support the hypothesis that vulnerable-to-bias decision procedures are less reliably executed than other decision procedures. Time available ranged from 45 seconds to 20 seconds. For all times available, the vulnerable-to-bias decision procedures were the less reliably executed procedures. This result occurred even though the decision procedures in which subjects were trained were *very* simple; involving only a few rules and no more than two preconditions per rule.

A possible confounding variable is the possibility that subjects adapted to time stress by routinely ignoring the other DMs advice. Although this behavior would consistent with the vulnerable-to-bias hypothesis, it would have the effect of reducing the effective sample size. Rather than treating each track as an independent sample, each set of tracks becomes a sample of one strategy shift. However, the results does not support this explanation. If true, this explanation would predict that performance for the BF,BS conditions would be the same as the BB,FF,SS condition, and that performance for the FB,FS and BS,SB condition would be near zero. This later result clearly did not occur.

Instead, the results suggest that subjects adapted to severe time stress by changing the order in which they processed tracks. For the lower stress conditions, subjects ordered tracks in the way they were trained; by first sending advice (DMj-merge), then taking care of the their own tracks (DMi-only), and finally by processing the tracks where they had to merge their judgment with the other DM's advice (DMi-merge). This was the optimal strategy since it (1) guaranteed that the other DM would have time to send advice, and (2) it left the most time consuming tracks to be processed last. Interestingly, as time stress increased, subjects adapted by moving away from the optimal strategy; processing tracks in the order DMj-merge, DMi-merge, and DMi-only. Indeed, this order was the worst possible order for processing tracks under time stress, since the first tracks to be completely processed (DMi-merge) are also the tracks that take the longest to process.

One possible explanation for this maladaptive behavior is that subjects responded to time stress by processing the tracks by "visual order", i.e., moving from upper left to the lower right portion of the screen. They were reducing cognitive workload by simplifying the decision procedure for selecting which track to process next.

Overall these results suggest that when specifying team decision procedures, vulnerable-to-bias procedures should be avoided. Although a well trained team may execute vulnerable-to-bias procedures reliably in non stress conditions, these procedures are very vulnerable to the effects of time stress. In addition, these results point to the importance of procedures for adapting to stress. The subjects in this experiment adapted inappropriately. Future experiments will examine the extent to which teams can be trained to adapt effectively, and whether adaptation strategies are themselves vulnerable-to-bias.

REFERENCES

Andreadakis, S.K. *Analysis and Synthesis of Decision-making Organizations*. Ph.D. Thesis, M.I.T., 1988.

Edwards, W, 1968, "Conservatism in human information processing" in Kleinmuntz, B, ed., *Formal Representation of Human Judgment* Wiley.

Einhorn, H and Hogarth, R, 1986, "Judging Probable cause" *Psychological Bulletin* **99** 3-19.

Fischhoff, B, 1975, "Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty" *Journal of Experimental Psychology:* Human perception and Performance 4 330-344.

Jin, V.Y. *On the effect of Organization Structure on Team Performance*. Ph.D. Thesis, M.I.T., 1990.

Kahneman, D and Tversky, A, 1973, "On the psychology of prediction" *Psychological Review* **80** 237-251.

Kahneman, D., Slovic, P. and Tversky, A. *Judgment under uncertainty: Heuristics and biases*.. Cambridge University Press, 1982.

Lehner, P.E. and Adelman, L. "Behavioral decision theory and its implications for knowledge engineering," *Knowledge Engineering Review*, 1990.

Levis, A.H. "Information processing and decision making organizations: A mathematical description." *Large Scale Systems*, 1984, 7, 1511-63.

Levis, A.H. and Boettcher, K.L. Modeling and analysis of teams of interacting decision makers with bounded rationality. *Automatica*, 1983, 9(6), 703-709.

Newell, A and Simon, H, 1972, *Human Problem Solving* Prentice Hall.

Nisbett, R and Ross, L, 1980, *Human Interface: Strategies and shortcomings of Social Judgment* Prentice-Hall.

Tversky, A and Kahneman, D, 1971, "The belief in the 'Law of Small Numbers'" *Psychological Bulletin* **76** 105-110.

Tversky, A and Kahneman, D, 1973, "Availability:A heuristic for judging frequency and probability" *Cognitive Psychology* **5** 207-232

Tolcott, MA, Marvin, FF and Lehner, PE, 1989, "Expert Decision making in evolving situations" *IEEE Transactions on Systems, Man and Cybernetics* **19**(3) 606-615.

Wason, P, 1960, "On the failure to eliminate hypotheses in a conceptual task" *Quarterly Journal of Experimental Psychology* **12** 129-140