AD643312

Technical Research Report 1151

AD

THE USE OF TEAMS IN IMAGE INTERPRETATION: INFORMATION EXCHANGE, CONFIDENCE, AND RESOLVING DISAGREEMENTS

by George W. Doten and John T. Cockrell,
System Development Corporation

and Robert Sadacca,
U. S. Army Personnel Research Office

D D C
DEC 14 1966

CLEARINGHOUSE
FOR FEDERAL SCIENTIFIC AND
$3.00 65 55 as
ARCHIVE COPY

OCTOBER 1966

Distribution of this document is unlimited

U. S. ARMY PERSONNEL RESEARCH OFFICE

U. S. ARMY PERSONNEL RESEARCH OFFICE

An activity of the Chief, Research and Development

ACCESSION for
DISTRIBUTION/AVAILABILITY CODES
DIST. AVAIL. and/or SPECIAL

J. E. UHLANER
Director
USAPRO Laboratories

M. O. BECKER
Colonel, GS
Commanding

Research accomplished
under Contract to the Department of the Army

SYSTEM DEVELOPMENT CORPORATION
Advanced Systems Division

NOTICES

DDC AVAILABILITY: Qualified requestors may obtain copies of this report directly from DDC. Available, for sale to the public, from the Clearinghouse for Federal Scientific and Technical Information, Department of Commerce, Springfield, Virginia 22151.

DISTRIBUTION: Primary distribution of this report has been made by APRO. Please address correspondence concerning distribution of reports to: U. S. Army Personnel Research Office, Washington, D. C. 20315. Copies may also be obtained on loan from local depository libraries. A list of these libraries may be obtained from: Documents Expediting Project, Library of Congress, Washington, D. C. 20540

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U. S. Army Personnel Research Office.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

AD

Technical Research Report 1151

THE USE OF TEAMS IN IMAGE INTERPRETATION: INFORMATION EXCHANGE, CONFIDENCE, AND RESOLVING DISAGREEMENTS

by George W. Doten and John T. Cockrell,
System Development Corporation
and
Robert Sadacca,
U. S. Army Personnel Research Office

SUPPORT SYSTEMS RESEARCH LABORATORY
Joseph Zeidner, Chief

U. S. ARMY PERSONNEL RESEARCH OFFICE

Office, Chief Research and Development
Department of the Army

Washington, D. C. 20315

October 1966

Army Project Number
2J620901A721

Contract No. DA-49-092-ARO-65
Component Integration Task

Distribution of this document is unlimited

USAPRO Technical Research Reports and Technical Research Notes are intended for sponsors of R&D tasks and other research and military agencies. Any findings ready for implementation at the time of publication are presented in the latter part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

The SURVEILLANCE SYSTEMS research program of the U. S. Army Personnel Research Office has as its objective the production of scientific data bearing on the extraction of information from surveillance displays, and the efficient storage, retrieval, and transmission of this information within an advanced computerized image interpretation facility. Research results are used in future systems design and in the development of enhanced techniques for all phases of the interpretation process within the data reduction facility. Research is conducted under Army RDT&E Project No. 2J620901A721, "Surveillance Systems: Ground Surveillance and Target Acquisition Interpreter Techniques," FY 66 Work Program.

U. S. APRO research under this Project is conducted as an integrated in-house and contractual effort, the latter provided by organizations selected as having unique capabilities and facilities for research in aerial surveillance. The Component Integration Task is one of four research Tasks established to focus on operationally meaningful segments of the surveillance system. Among the specific objectives of the Task is the identification of effective team procedures under various system conditions and requirements.

The present study was conducted jointly by personnel of the Advanced Systems Division, System Development Corporation and of the U. S. Army Personnel Research Office and centers on system team interactions designed to reduce the time required for team interpretation while maintaining the superiority of team procedures in the accuracy and completeness of the information extracted from imagery. The study was performed under the technical direction of Dr. Robert Sadacca, USAPRO, who is also a co-author. In addition to Dr. Sadacca, valuable comments and suggestions were received from Dr. John Mellinger, USAPRO.

J. E. UHLANER
Director
USAPRO Laboratories

THE USE OF TEAMS IN IMAGE INTERPRETATION: INFORMATIC ' EXCHANGE, CONFIDENCE, AND RESOLVING DISAGREEMENTS

BRIEF

Requirement:

To determine which aspects of image interpreter team operation are important in decreasing the amount of time required for team interpretation while maintaining the superiority of teams in accuracy and completeness. A secondary requirement was to investigate various methods of team operation.

Procedure:

Using the common procedure of having each team member in two-man teams check the interpretations of his teammate, three experiments centered around the following questions: (1) How much knowledge should the checker have of the initial interpreter's work? (2) How accurately can the initial interpreter rate the accuracy of his interpretations and can the initial interpreter effectively designate which of his interpretations need checking? And (3), how can a third interpreter best be utilized to resolve conflicts in interpretations made by the original two-man team? Variations centered about the amount of information passed from initial interpreter to checker, discussion between team members versus no discussion, consensus versus one-man decision in determining the team product, confidence ratings made by interpreters and confidence levels below which interpretations were checked, and participation of a third team member under varying conditions to resolve conflicts in interpretation. Results were evaluated in terms of completeness of information extracted, total amount of error, accuracy, and efficiency.

Findings:

1. Teams in which the checker had complete knowledge of the initial interpreter's work produced more complete results with higher efficiency.

2. Initial interpreters can judge only to a limited extent the adequacy of their interpretations. Using judgments as a means of limiting the amount of checking increased efficiency and did not appreciably affect accuracy or completeness. However, these results were somewhat ambiguous and definite conclusions should not be drawn at this time.

3. Introduction of a third man provided more completeness than the two-man team but reduced efficiency. There were no differences in team output resulting from different procedures with the three-man team.

4. Results with different team methods pose a tradeoff situation, since no one method can be considered best for team performance under all requirements. The checking procedure with arbitrary scoring resulted in the highest completeness but lowest accuracy. The checking procedure with consensus yielded higher accuracy but less complete interpretation. The discussion procedure with the consensus scoring gave both high accuracy and high completeness but reduced efficiency.

Utilization of Findings:

Based on tactical requirements, image interpreter team methods should reflect relative emphasis on completeness, accuracy, and efficiency. When complete information is required from an imagery mission, and timeliness is essential, team members should check each other's work without discussion, and decisions made by the checker should constitute the product. When a greater degree of accuracy is desired, only information agreed upon by the team members should be accepted. A reasonable balance between completeness, accuracy, and efficiency is achieved in two-man teams by adding the discussion procedure and then accepting only information agreed upon by the team members. Although not tested directly, the data also suggest that a reasonable compromise method would be to omit the discussion and use a third man to resolve conflicts, provided the consensus scoring rule was used. In all cases above, the checker should have full knowledge of the initial interpreter's work.

THE USE OF TEAMS IN IMAGE INTERPRETATION: INFORMATION EXCHANGE, CONFIDENCE, AND RESOLVING DISAGREEMENTS

CONTENTS

TABLES

TABLES (Continued) Page

FIGURES

THE USE OF TEAMS IN IMAGE INTERPRETATION: INFORMATION EXCHANGE, CONFIDENCE, AND RESOLVING DISAGREEMENTS

Research studies on the use of teams for image interpretation conducted at the U. S. Army Personnel Research Office (U. S. APRO) during the last several years have focused on the basic question of whether teams can perform image interpretation more effectively than can individuals acting alone, and on related questions concerning the best team methods and procedures and the best size of teams for maximizing performance.

The first of these studies demonstrated that teams of interpreters can extract more accurate and more complete information from imagery than can individuals. A second study demonstrated that gains in accuracy and completeness vary with team organization, size, and work procedures, also that teams are of the most value in interpreting relatively difficult imagery. A subsequent pilot study in which size of teams, amount of checking, and team organization were varied supported the superiority of teams, particularly in handling more difficult imagery.

These studies advanced the knowledge of team performance in image interpretation to the point where the effect of using teams should be considered in relation to the total amount of interpreter time spent on a given interpreter mission. While precise relationships between time, accuracy, and completeness have not been established for team performance, the evidence available indicates that, to process a given amount of imagery, teams require more man hours and possibly more total elapsed time than do individual interpreters. Moreover, no particular team method appears superior for all missions. The findings suggest rather that team procedures should be varied to meet specific mission requirements.

The principal research problem concerning the use of teams in image interpretation, then, is not whether team reports are more accurate and complete than individual reports, but when--that is, under what conditions--teams should be formed and how teams should operate to meet specific performance requirements, particularly with regard to timeliness. Consider that teams working within image interpretation facilities must be able to shift from rapidly processing large amounts of imagery to processing small amounts of imagery is a more detailed manner. Consider also that requirements may shift from the demand for very high accuracy to the demand for very high completeness. The quality of the imagery under operational conditions will probably vary considerably. These shifting circumstances necessitate the development of interpretation procedures which teams can employ flexibly to maximize accuracy, completeness, or timeliness according to the requirements that are levied.

SPECIFIC OBJECTIVES OF THE PRESENT STUDIES

The general purpose of the present set of experiments was to investigate those aspects of team operation which may result in a decrease in the time required for team interpretation while maintaining the superiority of teams in the accuracy and completeness of the information extracted. This basic purpose was translated into the following three specific primary objectives:

1. To determine the amount and type of knowledge which the checker should have of the initial interpreter's work.

2. To determine whether the initial interpreter can accurately determine when his work needs to be checked by his teammate.

3. To determine how best to utilize a third man to resolve disagreements among teammates on items of interpretation.

The following objectives were secondary:

1. To determine whether performance varies with the aptitude and proficiency of members of the team.

2. To compare the usefulness of various team methods combining different procedures and different means of combining the output of individual team members into a team output (scoring rules). A team method consisted of a team procedure plus a scoring rule.

FRAMEWORK FOR THE EXPERIMENTATION

Variations in procedures were achieved for analysis by setting up four phases or modules of interpreter team activity. The output from each module when combined with a scoring rule could be considered the final team product if interpretation were stopped with the completion of the module. Modules were so designed that as teams went through the four modules, members interacted more and more.

Module 1, the initial interpretation phase, involved almost no interaction among teammates. For each mission, interpreters worked independently on separate parts of the imagery.

Module 2 was the checking phase. During this phase the checker was provided knowledge of the initial interpreter's work to varying degrees. The checker's job was to check his teammate's identifications and look for and identify undetected targets. The interaction did not involve discussion. There were several possible results of the checking process: (1) The checker could agree with his teammate's identifications. (2) The checker could disagree with his teammate's identification either by identifying the object in question as a different target or by denying that the object was a target of military significance. Or (3) the checker could identify objects on the imagery that the first interpreter had omitted either deliberately or inadvertently. This action was also considered a disagreement.

Modules 3 and 4 were introduced in order to evaluate procedures for resolving disagreements between two team members. Module 3 was a discussion phase in which teammates considered conflicts in interpretation, exchanging ideas and reasons which had led them to a particular interpretation. In Module 4, a third interpreter joined the team to attempt to resolve conflicts either before discussion or after discussion. The third man did not discuss the disagreements with the original team members and he did not always have knowledge of the original identifications.

A number of team methods were set up incorporating variations in team procedures and means of combining individual interpretations into a final team interpretation (scoring rules). All interpreters worked through Module 1 which was common to all experimental procedures. Modules 2, 3, and 4 were experimental procedures which were combined with different scoring rules resulting in the various team methods. Scoring rules for team output centered about the emphasis given to the target interpretations made by the checker. In two-man teams where each interpreter checked the work of the other, the alternatives were (1) to accept only identifications which the two members agreed upon (consensus) and (2) to accept identifications which the checker verified or added, eliminating only those identifications rejected by the checker (arbitrary decision). The consensus scoring rule was applied in two team methods: consensus plus checking (Module 2) and consensus plus discussion (Module 3). The arbitrary scoring rule was applied in two similar team methods. When a third interpreter was introduced to resolve disagreements occurring in two-man teams, four additional team methods were formed, based on whether the third man was introduced before or after the original team members discussed their conflicts and on whether the consensus or arbitrary scoring rule was applied to the third man's work.

The team product achieved by each method was assessed in terms of completeness, amount of error, accuracy, and efficiency.

The imagery used in the experiments consisted of aerial photographs of Army field maneuvers. The imagery was subdivided into missions. In the initial phase, each interpreter processed approximately half the imagery of the mission assigned to the team. Each team member had his own viewing device--a light table--and other basic interpreter equipment. Processing consisted of searching each frame for designated types of military targets, annotating the imagery by circling and numbering the targets, and then identifying the targets, writing the number and identification on a report form. Each frame was processed completely before the next frame was started, and the interpreter proceeded without interruption through the half mission. Time limits for the different phases were set so as to rush the interpreters slightly, but to allow time for completion.

Image interpreter trainees about to graduate from the image interpretation course at the U. S. Army Intelligence School at Fort Holabird, Maryland, participated as subjects in the studies. Thirty-two interpreters were used in each of the first two experiments, 36 in the third experiment. There was some overlap in the subjects performing in the three experiments.

EXPERIMENT I. EFFECT OF INFORMATION EXCHANGED BY TEAM MEMBERS

In the first of the three experiments conducted, the checker had varying degrees of knowledge of the initial interpreter's work. Only two-man teams were considered, and interpretation was stopped after the discussion period, Module 3. Four conditions of information exchange were established:

No knowledge. The checker knew only the number of targets his teammate had found on each frame of imagery. The checker received a list which showed frame number, number of annotations, and number of targets identified by the initial interpreter.

Annotations only. The checker was allowed to look at the imagery annotated by his teammate, but did not see the report form containing the target identifications.

Identifications only. The checker was allowed to see a list of the targets identified by his teammate for each frame, but did not know the location on the frame of the objects identified.

Complete knowledge. The checker saw both the annotated imagery and the list of target identifications made by his teammate; that is, he knew where on the frame his teammate had located targets and what he had called them.

The results of Experiment I showed that the complete knowledge condition produced the highest completeness and efficiency. There was no difference in accuracy or total error. Insofar as team methods are concerned, the use of different scoring rules following the checking procedure greatly influenced team output, leading to high accuracy and low completeness for the consensus rule and the reverse of this for the arbitrary rule. Adding the discussion procedure greatly helped the consensus rule, raising completeness and only slightly reducing accuracy. The discussion procedure had very little effect on the arbitrary rule, raising completeness slightly and producing no change in accuracy. Efficiency was reduced by the discussion. The net overall result for team methods is that no one method gave the highest score on all measures.

EXPERIMENT II. INTERPRETERS' CONFIDENCE RATINGS USED TO LIMIT AMOUNT OF CHECKING

The second experiment explored the use of confidence levels. The objective was to determine the usefulness of having each interpreter in a two-man team indicate how confident he was that his identification of each target was correct before submitting his report to the checker. Can an interpreter decide reliably which of his identifications need to be checked by a second interpreter? If he can, this ability to discriminate his "sure" from his "uncertain" identifications could be used to reduce the amount of checking done and thus save valuable time with minimal loss of accuracy and completeness.

A major problem in using confidence estimates to control team checking operations is to select the level of confidence above which no checking will be done and below which all interpretations will be checked. As this cutoff value will most certainly vary with intelligence requirements for speed, accuracy, and completeness, results were evaluated at several levels of confidence: 100 percent (meaning that all annotations and identifications were checked), 80 percent (all annotations and identifications under that level were checked), 60 percent, and 40 percent. Use of confidence levels was applied only with two-man teams. Both the two-man consensus rule for acceptance of the team product and the rule of arbitrary decision of the checker were applied at each of the confidence levels established. Along with confidence in identification of individual targets, the value of confidence in detection of targets was studied. After completing each frame, interpreters rated their confidence that they had detected all targets present in that frame.

The results of the second experiment indicated that image interpreters have only a marginal ability to judge whether their interpretations need to be checked or not.[1] In regard to determining whether they have found all the targets on a particular frame, the interpreters show practically no ability.[2] These results indicate that before confidence could be used as a signal for the need of a check, considerable training in making confidence judgments would probably be necessary.

Use of confidence judgments to signal the need for checking produced very little effect on team output. As might be expected, the efficiency of the group that did the least amount of checking (40% group) was highest. Efficiency was the only measure for which results were significant. For accurrcy, completeness, and total error, there were no

[1] The correlation between confidence and accuracy of interpretation was +.41. This correlation just misses being significant statistically.
[2] Correlation coefficient of +.12.

differences among the groups. Although these results might appear to indicate that the least amount of checking is the preferred procedure, there were several trends which contradict this.

For example, additional probing of the amount of checking which took place under varying confidence level requirements indicated that the amount of unnecessary checking (that is, the number of targets correctly identified by the first interpreter which were checked by the second) was reduced as the confidence level at which checking was required was lowered. But, unfortunately, so was the amount of necessary checking (wrong identifications that should have been checked but were not).

Another example was the clear-cut trend for the group that did the most checking (100% group) to eliminate the most errors, to find and identify the most new targets, and to make the most additional errors. The conclusion is that the use of confidence ratings made by untrained interpreters is not a reliable technique to signal the need for checking.

One further finding from Experiment II was that interpreters tended to lower their confidence ratings as the cutoff level was reduced. They may have deliberately lowered their ratings knowing that only identifications and target detections below the cutoff level would be checked. To the extent that interpreters adjust their confidence estimates downward, the purpose of using confidence levels to reduce the amount of checking is defeated.

EXPERIMENT III. INTRODUCTION OF A THIRD INTERPRETER TO RESOLVE DISAGREEMENTS

The third experiment concentrated on ways of resolving disagreements in interpretations produced by two-man teams. The question was whether introduction of a third man could revolve disagreements in such a way as to improve the team product. Disagreements included identifications unique to either the original interpreter or the checker of the two-man team as well as identifications on which the two teammates were at variance. The third man directed his attention entirely to items of disagreement and did not look for additional targets. Three modes of resolving disagreements were studied:

1. The third interpreter attempted to identify all targets about which the other two team members were not in agreement at the end of the checking phase (module 2) but prior to the discussion phase (module 3). He had available the annotated imagery of the two-man team but not the identifications.

2. This mode differed from the first only in the amount of knowledge the third man had concerning interpretations already made. He had available the target identifications made by members of the two-man team as well as their annotations.

3. In this mode, the third man entered the team operation following the discussion phase. He resolved only those conflicts which remained after discussion. He had full information on the product of the two-man team--both annotations and target identifications, as well as the results of the discussion phase.

The different modes of resolving conflicts did not yield significantly different results on any measures of team performance. Results on modes of resolution were similar whether the scoring rule for the team product was a consensus of the third man with one or both members of the original team or the third man's decision on all disputed identifications.

Further comparisons of interest were between two-man and three-man teams and between scoring rules. The three-man teams had a higher completeness score than the two-man teams and a smaller number of errors. No differences were found in the accuracy measure which reflected correct interpretations in relation to total interpretations made. The two-man teams were more efficient, producing more correct identifications per unit of time spent.

Arbitrary decision by the checker in the two-man team and by the third interpreter in the three-man team produced significantly better performance on completeness, total error, and efficiency. The consensus standard led to greater accuracy of team output. These results were consistent with those obtained in the first two experiments.

DIFFERENCES IN TEAM COMPOSITION

A secondary purpose of the study was to note any variations in team performance associated with differences in the composition of the teams. For each of the first two experiments, 16 two-man teams were formed using General Technical (GT) Aptitude Area[3] scores and grades in the image interpreter course to identify individuals for assignment to teams characterized as high-high, high-low, medium-medium, and low-low. For Experiment III, interpreters were randomly assigned to the teams.

No significant differences were found among teams differing in composition on the basis of ability, either in Experiment I when the amount of information exchanged was varied or in Experiment II in which the use of confidence ratings was investigated. The implication is that aptitude scores and course grades are not effective predictors of interpreters' contributions to team output.

---

[3] A composite score on two tests of the Army Classification Battery--Verbal and Arithmetic Reasoning.

IMPLICATIONS OF THE FINDINGS

The results obtained in the three experiments support the tentative conclusion that it is possible to maintain team superiority while improving the timeliness of the team output.

In Experiment I, the condition affording the checker the most knowledge of the initial interpreter's work yielded the highest completeness and efficiency scores. In Experiment II, the highest efficiency score was obtained when only those interpretations with a confidence rating of 40% or less were checked, with no loss in accuracy, completeness, or increase in total error. While the results of Experiment II were not conclusive, they are encouraging in that they point to the feasibility of reducing unnecessary checking. The third experiment, testing the advantage of adding a third man to resolve conflicts in interpretation, was inconclusive. However, the general trends reinforced those noted in the first two experiments. The results with regard to team methods pose a dilemma since no one team method resulted in the highest score on all team measures. Completeness was highest when the checking procedure was used with the arbitrary scoring rule. Accuracy was highest when the checking procedure was used with the consensus scoring rule. The discussion procedure used with the consensus scoring rule was the best compromise between accuracy and completeness, but unfortunately efficiency was lowered. Adding a third man after the checking procedure and using the consensus rule was also an effective compromise between accuracy and completeness; however, this method also reduced efficiency.

The implications of the findings suggest that team methods must be tailored to meet mission requirements, and that no one method will be best for all team scores. The user must choose between completeness and accuracy or be content with reduced efficiency.

The team methods which have been used so far do not exhaust the possible methods which could be used with teams. Three possible approaches to the problem of increasing all team scores were suggested by the outcome of the studies:

1. Instruct the initial interpreter to strive for completeness rather than accuracy. This approach stems from the fact that checkers seem to be able to correct the errors made by the initial interpreter, but harm the team product mostly by adding errors of their own.

2. Train the initial interpreter to make more exact confidence ratings so that the ratings are a more reliable signal for the need of a check.

3. A completely different approach to the problem would be to select interpreters according to their ability to perform the different aspects of the job. Before this approach could be taken, it would be necessary to determine if interpreters have any differential ability to perform the various tasks. Teams formed to take advantage of any differential abilities so detected could be compared with teams formed at random.

THE USE OF TEAMS IN IMAGE INTERPRETATION: INFORMATION EXCHANGE, CONFIDENCE, AND RESOLVING DISAGREEMENTS

TECHNICAL SUPPLEMENT

TECHNICAL SUPPLEMENT

EXPERIMENTAL DESIGN AND RESULTS OF THREE EXPERIMENTS ON IMAGE INTERPRETER TEAM METHODS

The three experiments conducted were each directed toward one of the three primary objectives of the present study. Certain analyses were replicated in two or all the experiments, particularly analyses concerned with team methods.

Certain methodological elements were common to the three experiments.

Team Procedures

The team procedures used in the study consisted of four modules or subsets of activities as described in the test of the report. The modules are noted briefly below:

Module 1. Initial interpretation. Members of two-man teams worked independently on separate parts of the imagery, completing annotations and target identifications.

Module 2. Checking. Teammates checked each other's interpretations and looked for additional targets.

Module 3. Discussion. Teammates discussed identifications on which they did not agree.

Module 4. Use of Third Man. A new member joined the team. He checked the interpretations on which the original team members had not reached agreement.

Team Scoring Rules

A scoring rule was defined as a means of combining individual output into a team output. The two basic scoring rules were:

Consensus. Score only responses which two teammates agree upon.

Arbitrary. Score all responses which checkers approve or make.

When a third man entered the team, the scoring rules were basically the same but slightly different in application, as follows:

Third man final (Arbitrary). Score any responses made by the third man and add them to the agreed upon responses produced in Modules 1 and 2.

Consensus (Two out of three). Add the third man's responses to the responses produced in Modules 1 and 2 only if the third man agrees with at least one of the first two men.

These team scoring rules are clearly differentiated from the team procedures. The word "procedure" here defines the subsets of activities or modules which made up the team operations. The scoring rules were applied to the product of the team operation. Together, a team procedure and a scoring rule constituted a team method in these experiments.

Team Methods

The basic scoring rules were applied to the team product that resulted before and after the discussion procedure (Module 3). Four team methods were therefore employed with two-man teams:

1. Checker pre-discussion. The team product was considered to comprise all responses made or approved by the checker without any discussion.

2. Two-man agreement pre-discussion. Only responses which the two teammates agreed upon prior to discussion were considered the team product.

3. Checker post-discussion. All responses still approved by the checker after discussion were considered the team product.

4. Two-man agreement post-discussion. Only responses which the two teammates agreed upon after discussion were considered the team product.

When a third man was employed, four additional team methods resulted. These are described in connection with Experiment III.

Dependent Variables

Four measures of team performance were used:

1. Accuracy. Ratio of right interpretations to the sum of right plus wrong interpretations.

2. Completeness. Ratio of right interpretations to the total possible rights, that is, the total number of scored targets in the imagery.

3. Total Error. Sum of three different kinds of wrong interpretations:

Inventive errors: the interpreter identified a non-military object as a target.

Misidentifications: the interpreter identified a target wrongly, e.g., identified a tank as a truck.

Errors of omission: interpreter failed to respond at all to a scorable target.

The total error score, in effect, weighted these three kinds of errors equally.

4. Efficiency. Number of right interpretations divided by the total amount of time required by the team.

Experimental Subjects

Image interpreter trainees about to graduate from the interpretation course at Fort Holabird, Maryland, constituted the population of subjects for the three experiments.

Experimental Imagery

The imagery used in the experiments consisted of aerial photographs taken of Army field maneuvers, subdivided into missions. All missions had the following characteristics:

Positive transparency roll film

9" x 9" size

Approximately 40-60% stereo overlap

Scale from 1:2000 to 1:5000

Approximately 24 photographs (frames) to each mission

23-76 military objects (targets) on each mission

0-15 targets on any one frame

EXPERIMENT I. EFFECT OF KNOWLEDGE CONDITIONS

Experimental Objectives

The primary objective of the first experiment was to determine the effect of different knowledge conditions on team performance. Four such conditions were selected:

Condition A: No knowledge. The checker knew only the number of targets his teammate had found on each frame. The checker was passed a list showing frame number, number of annotations, and number of targets.

Condition B: Annotated Imagery Only. The checker was allowed to look at his teammate's annotated imagery but was not allowed to look at the report form containing the identifications.

Condition C: Identifications Only. The checker was allowed to see his teammate's identifications, but did not know where on the frames the targets had been located by his teammate.

Condition D: Complete Knowledge. A combination of conditions B and C. The checker knew where his teammate had located targets and what he had called them.

Secondary objectives were to determine how performance varied as a function of the proficiency of individual team members and to compare various team methods.

Experimental Design

In order to balance knowledge conditions and missions, a replicated 4 x 4 x 4 x 4 Graeco-Latin square was used. There were four knowledge conditions (A, B, C, D), four missions (a, b, c, d), four test periods (I, II, III, IV), and four teams (1, 2, 3, 4) as shown below:

	Test Periods			
Teams	I	II	III	IV
1	Aa	Bb	Cc	Dd
2	Dc	Cd	Ba	Ab
3	Bd	Ac	Db	Ca
4	Cb	Da	Ad	Bc

This square was replicated four times, each square utilizing teams with different proficiency levels:

Square 1: 4 high-high teams (both teammates high in proficiency)

Square 2: 4 high-low teams (one teammate high, one low in proficiency)

Square 3: 4 medium-medium teams (both teammates medium in proficiency)

Square 4: 4 low-low teams (both teammates low in proficiency).

Thirty-two enlisted men from two image interpreter classes were used to form the 16 teams participating.

Team Procedures

Each team first went through the initial interpretation phase (Module 1). A 30-minute time limit was set for each mission. The initial interpreter was required to fill out, in addition to the standard target identification form, a form which indicated the number of targets found on each frame. This form was used in the checking phase described above for knowledge Condition A.

After the teams had finished the initial interpretation phase, the checking phase (Module 2) started immediately. Depending on the condition that a team was to enter, teammates either exchanged forms or exchanged seats or both, and began checking each other's work. At the end of this 30-minute phase, grid locations for each annotation number were entered on the report forms. Annotations and identifications were then compared and sorted as to agreement or disagreement. A 30-minute period was then allowed for teammates to discuss their disagreements (Module 3). If they eventually agreed upon an identification, they wrote it in the appropriate space on the checker's report form. The total time for a test period was the time required to complete the three modules. The forms given to the teams are reproduced in the Appendix.

Team Methods

The four basic team methods resulting from the application of two scoring rules--checker and two-man agreement--prior to and after the discussion module were employed: checker pre-discussion; two-man agreement pre-discussion; checker post-discussion; and two-man agreement post-discussion.

Results

The effect of the four knowledge conditions on the performance variables is shown in Table 1 which presents the mean accuracy, completeness, total error, and efficiency scores of the teams. Table 1 scores are the result of using the two-man agreement post-discussion method.

Condition D, the full information condition, produced significantly higher completeness and efficiency than the other conditions. The accuracy and total error scores did not vary to any great extent across the information conditions. Similar results were obtained for the other three methods, as shown in Tables 2, 3, 4, 5, and 6. These tables include the F ratios, mean squares, and sources of variation for the four performance variables.

Table 1

MEAN PERFORMANCE SCORES FOR KNOWLEDGE CONDITIONS UNDER THE TWO-MAN AGREEMENT POST-DISCUSSION METHOD (Experiment I)

Knowledge Condition	Accuracy	Completeness*	Total Error	Efficiency**
A. No Knowledge	88%	63%	16	.20
B. Annotations Only	87%	64%	15	.22
C. Identifications Only	89%	67%	15	.20
D. Complete Knowledge	85%	70%	15	.23

*Means significantly different, $P < .05$
**Means significantly different; $P < .01$

Table 2

MEAN PERFORMANCE SCORES FOR FOUR KNOWLEDGE CONDITIONS UNDER THREE TEAM METHODS (Experiment I)

Checking Pre-Discussion Method

Information Condition	Accuracy	Completeness**	Total Error	Efficiency**
A. No Knowledge	82%	63%	17	.22
B. Annotations Only	85%	65%	15	.25
C. Identifications Only	82%	64%	17	.22
D. Complete Knowledge	82%	70%	16	.25

Two-Man Agreement Pre-Discussion Method

Information Condition	Accuracy	Completeness**	Total Error	Efficiency**
A. No Knowledge	95%	56%	18	.19
B. Annotations Only	92%	57%	17	.21
C. Identifications Only	94%	55%	18	.18
D. Complete Knowledge	86%	63%	16	.22

Checking Post-Discussion Method

Information Condition	Accuracy	Completeness*	Total Error	Efficiency**
A. No Knowledge	82%	65%	18	.16
B. Annotations Only	84%	66%	15	.18
C. Identifications Only	83%	68%	16	.17
D. Complete Knowledge	84%	71%	15	.19

*Means significantly different, $P < .05$
**Means significantly different, $P < .01$

Table 3

SOURCE OF VARIATION, MEAN SQUARES, AND F-RATIOS FOR ACCURACY
OF IDENTIFICATIONS UNDER FOUR TEAM METHODS
(Experiment I)

Source of Variation	d.f.	Pre-Check	Post-Check	Pre 2-Man	Post 2-Man
Team Type (T)	3	2.86	2.15	.35	2.96
Teams Within Type (Mean Square)	12	.018	.018	.045	.0063
Periods (P)	3	1.98	1.68	.31	2.21
P x T	9	.39	.95	.80	.39
Information Conditions (IC)	3	.31	.95	.98	1.14
IC x T	9	.43	1.14	.57	.81
Missions (M)	3	5.88*	8.71**	.83	5.84*
M x T	9	1.67	.95	.50	2.94*
Mean Square (Residual Error)	12	.0095	.0080	.044	.0042

*Means significantly different, P < .05
**Means significantly different, P < .01

Table 4

SOURCE OF VARIATION, MEAN SQUARES, AND F-RATIOS FOR COMPLETENESS
OF IDENTIFICATIONS UNDER FOUR TEAM METHODS
(Experiment I)

Source of Variation	d.f.	Pre-Check	Post-Check	Pre 2-Man	Post 2-Man
Team Type (T)	3	0.72	1.13	0.78	1.24
Team Within Type (Mean Square)	12	.028	.025	.033	.024
Periods (P)	3	1.89	4.95*	2.42	9.57**
P x T	9	2.78	7.48**	2.14	12.38**
Information Conditions (IC)	3	4.29*	7.79**	3.65*	17.35*
IC x T	9	0.12	0.56	0.86	2.70
Missions (M)	3	101.78**	235.36**	72.11**	476.81**
M x T	9	1.01	2.34	0.70	3.88*
Mean Square (Residual Error)	12	.049	.0016	.0057	.00082

*Significant at .05 level
**Significant at .01 level

Table 5

SOURCE OF VARIATION, MEAN SQUARES, AND F-RATIOS FOR TOTAL ERROR IN IDENTIFICATIONS UNDER FOUR TEAM METHODS (Experiment I)

Source of Variation	d.f.	Pre-Check	Post-Check	Pre 2-Man	Post 2-Man
Team Type (T)	3	2.65	2.59	1.65	2.09
Teams Within Type (Mean Square)	12	38.85	38.53	41.03	31.84
Periods (P)	3	2.51	4.10*	1.64	4.79*
P x T	9	1.51	3.19*	1.02	3.01*
Information Conditions (IC)	3	1.51	4.02*	.63	.69
IC x T	9	.64	2.47	.65	1.94
Missions (M)	3	70.53**	129.40**	100.89**	208.91*
M x T	9	1.21	1.84	.85	2.91
Mean Square (Residual Error)	12	14.27	7.19	15.98	4.84

*Means significantly different, $P < .05$
**Means significantly different, $P < .01$

Table 6

SOURCE OF VARIATION, MEAN SQUARES, AND F-RATIOS FOR EFFICIENCY FOR FOUR TEAM METHODS (Experiment I)

Source of Variation	d.f.	Pre-Check	Post-Check	Pre 2-Man	Post 2-Man
Team Type (T)	3	1.09	1.76	1.46	1.52
Teams Within Type (Mean Square)	12	.0057	.0032	.0054	.0035
Periods (P)	3	3.50*	10.24**	1.41	10.38**
P x T	9	1.91	2.64	1.00	2.81*
Information Conditions (IC)	3	5.04*	6.60**	4.65*	5.92*
IC x T	9	.42	.83	.71	.63
Missions (M)	3	216.98**	212.75**	104.67**	214.93**
M x T	9	1.51	1.66	1.28	1.44
Mean Square (Residual Error)	12	.0011	.00046	.0013	.00051

*Means significantly different, $P < .05$
**Means significantly different, $P < .01$

Table 7 shows the mean performance scores obtained for the four team methods. Values were obtained by summing team identifications across missions disregarding knowledge conditions and test periods. The scores obtained for the four team methods were considered replication scores in the analysis of variance (see Table 8). The analysis essentially compared pre-discussion performance with post-discussion performance and the checker scoring rule with the two-man agreement rule. That is, for the first part of the analysis the checker pre-discussion and two-man agreement pre-discussion scores were combined and compared with the combined checker post-discussion and two-man agreement post-discussion scores. In the second part of the analysis, checker pre- and post-discussion scores were combined and compared with two-man agreement pre- and post-discussion scores.

Results indicated that the discussion module significantly raised team completeness scores but lowered accuracy scores. Total error, however, was reduced. On the other hand, the two-man agreement methods resulted in significantly lower completeness than the checker methods. However, higher accuracy scores were obtained. Total error was not significantly different for the two methods. Efficiency was highest for the checker pre-discussion method. These results are consistent with expectations based upon previous experimentation. Adding the discussion module to the two-man agreement method appears to effect a reasonable compromise; a relatively large increase in completeness is obtained accompanied by a small drop in accuracy.

Table 7

MEAN PERFORMANCE SCORES FOR FOUR TEAM METHODS*
(Experiment I)

Team Method	Accuracy	Completeness	Total Error	Efficiency
Checker Pre-Discussion	84%	65%	65	.23
Two-Man Agreement (Pre-Discussion)	92%	57%	69	.20
Checker Post-Discussion	84%	67%	62	.20
Two-Man Agreement (Post-Discussion)	88%	65%	61	.20

*Pre-Discussion vs. Discussion and Two-Man Agreement vs. Checker significantly different ($P < .01$) for all variable comparisons except Total Error for the Checker vs. Two-Man Agreement comparison.

Table 8

SOURCE OF VARIATION, MEAN SQUARES, AND F-RATIOS
FOR COMPARISON AMONG TEAM METHODS
(Experiment I)

Source of Variation	d.f.	Completeness	Accuracy	Total Error	Efficiency
Team Type (T)	3	1.34	3.02	2.59	1.44
Team Within Type (Mean Square)	12	.0243	.0078	526.06	.00041
Checker vs 2-Man Agreement Score (CA)	1	94.71*	32.75*	.53	144.81*
T x CA	3	.44	2.54	1.56	1.80
Error (Mean Square)	12	.00046	.0017	30.15	.00021
Pre- vs Post-Discussion (PP)	1	101.72*	10.43*	41.06*	37.21*
T x PP	3	2.19	.51	.96	0.07
Error (Mean Square)	12	.00040	.00061	12.60	.000058
CA x PP	1	56.99*	34.41*	12.69*	55.82*
T x CA x PP	3	.039	1.28	.096	.130
Error (Mean Square)	12	.00032	.00029	8.69	.000081

*Means significantly different, $P < .01$

Table 9 presents the performance scores achieved by the different team types. Values were obtained by averaging scores across missions and using the two-man agreement post-discussion team method. (Table 10 gives the team-type results for the other three methods.) None of the performance differences in Table 9 are significantly different. This result is somewhat surprising; the high-high teams were expected to perform best. Course grades and aptitude scores may not be effective predictors of team performance. For the team methods employing discussion, significant interactions for completeness scores were obtained between team types and test periods (see Table 4). High-low proficiency teams showed a pronounced increase in completeness over time, whereas the high-high teams showed a drop in performance (see Figures 1 and 2). The high-high teams may have found discussion relatively unproductive, whereas discussion may have spurred the high-low teams to greater productivity.

Table 9

MEAN PERFORMANCE SCORES FOR TEAM TYPES UNDER THE TWO-MAN AGREEMENT POST-DISCUSSION METHOD (Experiment I)

Team Type	Accuracy	Completeness	Total Error	Efficiency
High-high	87%	67%	15	.20
High-low	91%	71%	12	.19
Medium-medium	83%	65%	17	.18
Low-low	88%	61%	17	.15

Table 10

MEAN PERFORMANCE SCORES FOR TEAM TYPES UNDER THREE TEAM METHODS (Experiment I)

Checking Pre-Discussion Method

Team Type	Accuracy	Completeness	Total Error	Efficiency
High-high	85%	68%	15	.22
High-low	88%	68%	14	.26
Medium-medimu	75%	65%	19	.23
Low-low	82%	61%	17	.21

Two-Man Agreement Pre-Discussion

Team Type	Accuracy	Completeness	Total Error	Efficiency
High-high	92%	61%	16	.22
High-low	96%	59%	15	.23
Medium-medium	91%	56%	19	.20
Low-low	90%	53%	19	.18

Checking Post-Discussion

Team Type	Accuracy	Completeness	Total Error	Efficiency
High-high	84%	68%	15	.19
High-low	89%	72%	13	.19
Medium-medium	75%	67%	18	.17
Low-low	84%	62%	17	.15

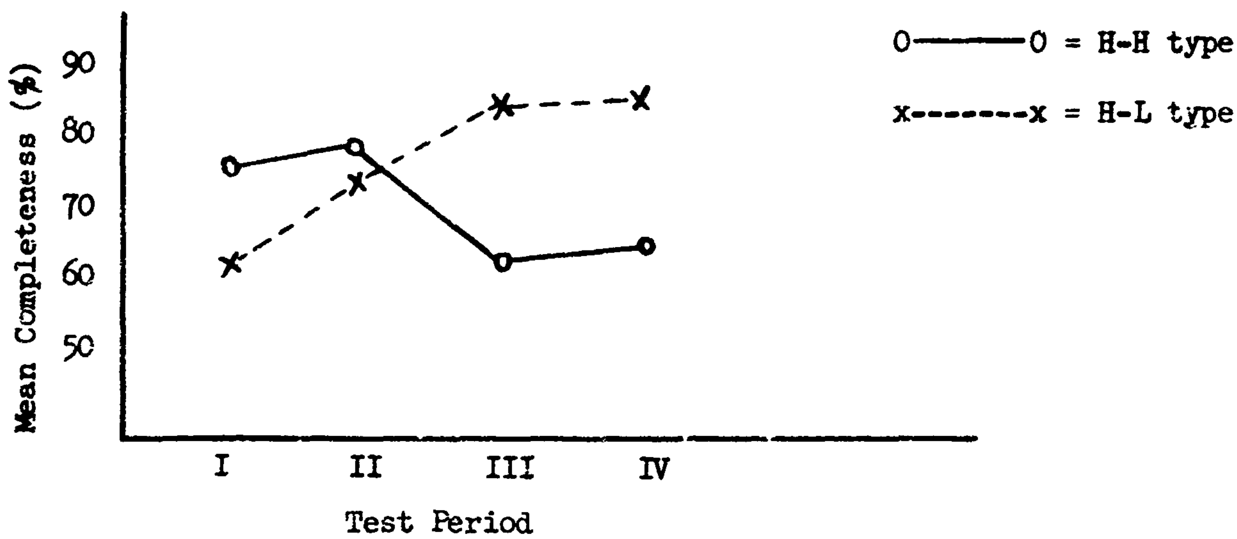

Figure 1. Completeness means for checker post-discussion method by test period

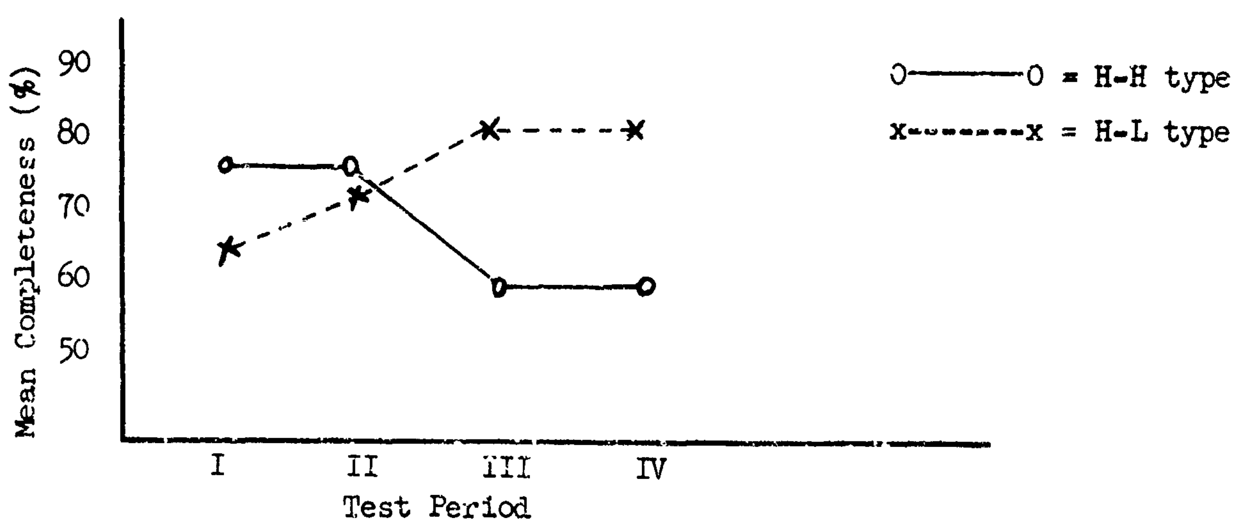

Figure 2. Completeness means for two-man agreement post-discussion method by test period

The missions employed in this experiment and in the other two experiments yielded significant differences in practically all analyses. This result reflects the large differences in difficulty of the imagery. Periods were significantly different in about half the analyses, performance improving with practice. Neither of these two variables--missions or periods--was considered of particular importance in the experiments except insofar as they showed evidence of interacting with the main experimental conditions. (Experiments studying team behavior over extended periods of time are planned in future U. S. APRO research.)

EXPERIMENT II. THE USE OF CONFIDENCE ESTIMATIONS IN CHECKING

Experimental Objectives

The primary purpose of the second experiment was to determine whether one teammate could accurately gauge when his interpretations needed checking by his teammate; and if so, whether this discrimination ability could be used to reduce the amount of checking, thus saving valuable time with minimal loss in accuracy and completeness. The assumption was: the more confident an interpreter, the less need there is to check his identification, and vice versa.

A major problem in utilizing confidence estimates to control team checking operations is to select the level of confidence above which no checking will take place and below which all interpretations will be checked. As this cutoff value would almost certainly vary as a function of the intelligence requirements for speed, accuracy, and completeness, four levels of confidence were used in this experiment:

Level A: 100%. A team member checked all his teammate's annotations and identifications. (In effect, confidence levels were not being utilized to determine checking behavior.)

Level B: 80%. A team member checked only the identifications and frames to which his teammate had assigned a confidence estimate of 80% or less.

Level C: 60%. A team member checked only identifications and frames to which his teammate had assigned a confidence estimate of % or less.

Level D: 40%. A team member checked only identifications and frames to which his teammate had assigned a confidence estimate of 40% or less.

These confidence levels constituted the main experimental factor of Experiment II. The confidence levels were applied against each identification made by the interpreters. In addition, the levels were applied against the interpreter's confidence that all targets on a frame had been detected. (After completing each frame, the interpreters rated their confidence that they had detected all targets on the frame.)

Experiment II had the same secondary objectives as Experiment I--to determine how performance varied as a function of proficiency and aptitude of individual team members and to compare various team methods.

Experimental Design

A replicated Graeco-Latin Square design, identical in all respects to the design used in Experiment I, was employed. The teams went through four imagery missions during four test periods, each time using a different level of confidence to determine checking behavior. The design included replications by the four team proficiency types: high-high, high-low, medium-medium, and low-low. A sample of 32 enlisted men from two image interpreter classes was used to form the 16 teams. Half of these were interpreters who participated in Experiment I.

Team Procedures

Each team went through the initial interpretation phase (Module 1). The individual interpreters recorded on their report form (See Appendix) their confidence in each identification immediately after making the response. In making their confidence estimations, the interpreters used a scale of 0-100%, with 100% indicating they were 100% positive they were correct, 90% indicating they felt they had an 90% chance of being correct, etc. After completing each frame, the interpreters similarly estimated their confidence that they had detected all targets on the frame. The team members were told beforehand what cutoff level would be used in the checking phase and therefore knew the operational implications of the confidence levels they assigned.

After the teams had finished the initial interpretation phase, the checking phase (Module 2) started immediately. Condition D of the first experiment was used in the checking phase. This condition allowed the checker to see both the identifications and annotations as well as the confidence levels of his teammate. Experiment II did not have a discussion phase.

Team Methods

As there was no discussion module, the application of two scoring rules--checker and two-man agreement--resulted in two team methods: checker pre-discussion and two-man agreement pre-discussion.

Results of Experiment II

The effect of the four confidence levels on the performance variables may be seen in Table 11 which shows the mean accuracy, completeness, total error, and efficiency scores of the teams for the checking and two-man agreement methods. None of the mean scores was significantly different among the confidence levels, with the exception of efficiency, which was highest for the 40% confidence level under both the checker and two-man agreement methods (see analysis of variance, Tables 12 and 13). Efficiency, a measure of the number of right responses produced per unit of time (minutes), was expected to increase as the number of responses to be checked decreased; the most timely and efficient procedure would, most probably, be to have no checking at all. However, as shown by previous experimentation[4], [5], [6], poorer accuracy and completeness would most probably also result.

Table 11

MEAN PERFORMANCE SCORES FOR LEVELS OF CONFIDENCE IN IDENTIFICATIONS APPLIED WITH CHECKER AND TWO-MAN AGREEMENT PRE-DISCUSSION METHODS (Experiment II)

Checker

Confidence Level	Accuracy	Completeness	Total Error	Efficiency*
100%	80	62	18	.23
80%	80	60	17	.23
60%	78	55	21	.23
40%	83	60	17	.28

Two-Man Agreement

Confidence Level	Accuracy	Completeness	Total Error	Efficiency**
100%	87	52	20	.18
80%	86	51	20	.19
60%	81	49	20	.21
40%	83	55	19	.26

*Mean values significantly different, $P < .05$
**Mean values significantly different, $P < .01$

[4] Sadacca, R., Martinek, H., and Schwartz, A. I. Image Interpretation Task--Status Report. USAPRO Technical Research Report 1129. Washington: U. S. Army Personnel Research Office, June 1962.

[5] Bolin, S. F., Sadacca, R., and Martinek, H. Team Procedures in Image Interpretation. USAPRO Technical Research Note 164. Washington: U. S. Army Personnel Research Office, December 1965.

[6] Bolin, S. F., Cockrell, J. T., and Doten, G. W. Basic Plan and Preliminary Results of the Photo Interpretation Team Studies Subtask. Unpublished pilot study. Washington: U. S. Army Personnel Research Office, March 1965.

Table 12

SOURCE OF VARIATION, MEAN SQUARES, AND F-RATIOS
FOR CHECKING TEAM METHOD
(Experiment II)

Source of Variation	d.f.	Total Error	Completeness	Accuracy	Efficiency
Team Type	3	2.57	2.11	.42	1.84
Team Within Type (Mean Square)	12	22.37	.015	.0190	.0080
Periods	3	2.05	1.86	2.69	5.09**
Periods x Team Type	9	1.07	.99	.87	.78
Amount of Checking	3	1.43	2.15	.80	3.52*
Amount of Checking x Team Type	9	1.19	.99	1.36	1.26
Missions	3	97.36**	60.14**	17.28**	45.50**
Missions x Team Type	9	3.45	1.73	1.40	.65
Residual Error (Mean Square)	12	12.29	.0076	.0068	.0033

*Means significantly different, $P < .05$
**Means significantly different, $P < .01$

Table 13

SOURCE OF VARIATION, MEAN SQUARES, AND F-RATIOS
FOR TWO-MAN AGREEMENT TEAM METHOD
(Experiment II)

Source of Variation	d.f.	Total Error	Completeness	Accuracy	Efficiency
Team Type	3	.70	1.22	.10	1.01
Teams Within Type (Mean Square)	12	30.24	.017	.014	.012
Periods	3	4.81*	5.69*	2.00	10.37*
Periods x Team Type	9	1.57	1.56	.75	1.03
Amount of Checking	3	.95	2.09	1.52	7.80*
Amount of Checking x Team Type	9	1.87	1.89	1.15	1.71
Missions	3	160.41*	101.05*	12.49*	57.42*
Missions x Team Type	9	2.40	1.75	.90	.99
Residual Error (Mean Square)	12	8.62	.0046	.0092	.0024

*Means significantly different, $P < .01$

Additional analyses were conducted to determine the effect of the different confidence levels on the checking activity of the teams. For this analysis, total response characteristics were examined across all teams and missions. Table 14 shows the total number of correct and incorrect identifications checked and not checked, as well as the mean amount of time required by each team for the checking phase. Table 15 shows the effect of employing the different confidence levels on the detection activity (the search for additional targets) of the checkers. The total number of frames with and without additional undetected targets that were checked and not checked is shown. The amount of time required for detection is intermingled with that required for identification and was not measured separately. Tables 14 and 15 indicate that the amount of unnecessary checking was reduced as the cutoff confidence level was lowered, but unfortunately so was the amount of necessary checking.

Table 14

TOTAL NUMBER OF RIGHT AND WRONG IDENTIFICATIONS CHECKED AND NOT CHECKED UNDER FOUR CONFIDENCE LEVELS ACROSS ALL TEAMS AND MISSIONS (Experiment II)

Confidence Level	Wrongs Checked	Wrongs Not Checked	Rights Checked	Rights Not Checked	Mean Time (Minutes)
100%	78	0	305	0	50
80%	64	11	115	182	45
60%	56	20	76	201	43
40%	36	30	51	260	40

Table 15

TOTAL NUMBER OF ADDITIONAL TARGET (AT) FRAMES AND NO ADDITIONAL TARGET (NO) FRAMES CHECKED AND NOT CHECKED UNDER FOUR CONFIDENCE LEVELS ACROSS ALL TEAMS AND MISSIONS (Experiment II)

Confidence Level	AT Frames Checked	AT Frames Not Checked	NO Frames Checked	NO Frames Not Checked
100%	115	0	229	0
80%	75	43	124	102
60%	65	63	96	120
40%	51	70	75	148

Table 16 shows the net effects on the number of right and wrong responses for both the identification and frame checks. The results are given in totals for all teams under each checking condition. From Table 16, identification checking is seen to have reduced errors and to have had very little effect on number of correct responses. The more checking, the more errors were eliminated. Detection checking, on the other hand, added both correct responses and incorrect responses. The more checking, the more responses of both types were added. The combined net effect of checking was no change for incorrect responses and the addition of many correct responses; the addition was greater the higher the cutoff confidence level employed.

Additional analysis was conducted to determine the effect on the confidence ratings themselves of the cutoff confidence levels employed in the experiment. Four variables were generated from the responses made by both team members to a given mission during Module 1, the independent phase:

1. Average confidence rating assigned to all identifications.

2. Average confidence rating assigned to each frame.

3. Validity of the identification confidence ratings (measured by the point biserial correlation between confidence and identification-accuracy).

4. Validity of the frame confidence ratings (measured by the point biserial correlation between confidence and the presence of additional undetected targets).

The results shown in Table 17 indicate that interpreters tended to lower their confidence ratings as the confidence cutoff level was lowered (see analysis of variance, Table 18). The interpreters may have deliberately lowered their ratings knowing that responses with confidences below the cutoff level would be checked. To the extent interpreters do adjust their confidences downward, the purpose of using lower cutoff levels to achieve greater checking timeliness is defeated. As indicated in Table 17, however, the validity of the confidence ratings was not significantly affected by the cutoff levels employed. The validity coefficients varied widely across the 16 teams; identification validity ranged from .18 to .62, frame validity from -.13 to .45. Across all teams and missions, an overall mean validity coefficient of .41 was obtained for the identification ratings. The frame rating mean coefficient was only .12. Although the identification validity is encouraging, considerable training in making confidence judgments would probably be necessary before such judgments would be sufficiently accurate and reliable for operational usage.

Table 16

NET EFFECT OF CONFIDENCE LEVELS ON NUMBER OF RIGHTS AND WRONGS IN CHECKING ACROSS ALL TEAMS AND MISSIONS (Experiment II)

	Confidence Level			
	100%	80%	60%	40%
Identification Checking				
Wrongs to Right	8	2	4	3
Wrongs Negated	32	24	18	11
Rights to Wrong	1	3	1	0
Rights Negated	3	6	0	1
Net Change for Identification				
Rights	+4	-7	+3	+2
Wrongs	-39	-23	-21	-14
Detection Checking				
Additional Rights	58	47	34	27
Additional Wrongs	35	32	24	13
Net Change for All Checking				
Rights	+62	+40	+37	+29
Wrongs	-4	+9	+3	-1

Table 17

MEAN CONFIDENCE RATINGS AND VALIDITY COEFFICIENTS OF CONFIDENCE RATINGS AT FOUR CONFIDENCE LEVELS (Experiment II)

Confidence Level	Mean Identification Rating*	Mean Frame Rating*	Identification Validity	Frame Validity
100%	78%	77%	.48	.07
80%	81%	78%	.47	.18
60%	72%	68%	.39	.09
40%	68%	57%	.29	.15

*Means significantly different, $P < .01$

Table 18

SOURCE OF VARIATION, MEAN SQUARES, AND F-RATIOS
FOR CONFIDENCE AND CORRELATION
(Experiment II)

Source of Variation	d.f.	Confidence		Correlation	
		ID	Detection	ID	Detection
Team Types	3	2.95	1.18	.04	.72
Teams Within Type (Mean Square)	12	.00015	.00044	.0000086	.000011
Periods	3	7.32*	.63	.38	1.67
Periods x Team Type	9	1.90	2.67	1.82	.62
Amount of Checking	3	35.66*	47.08*	1.54	.43
Amount of Checking x Team Type	9	2.46	2.53	.80	1.10
Missions	3	16.94*	6.16*	1.04	.40
Missions x Team Type	9	3.52*	2.27	.09	.41
Residual Error	12	.000016	.00003	.000012	.0000089

*Means significantly different, P < .01

Table 19 shows the mean performance scores obtained for the two team methods. Values were obtained by summing team identifications across missions, disregarding confidence levels and sessions. Since the values were similar in magnitude to those found in Experiment I, and the direction of differences was identical, no statistical tests of significance were performed. The checker method again produced higher completeness and efficiency rates and lower overall total error. The two-man agreement method, however, produced higher accuracy.

Table 19

MEAN PERFORMANCE SCORES FOR TEAM METHODS
(Experiment II)

Team Method	Accuracy	Completeness	Total Error	Efficiency
Checker Pre-Discussion	80%	59%	73	.24
Two-Man Agreement (Pre-Discussion)	84%	51%	78	.21

The analysis of variance did not reveal any significant differences in performance scores among the different team types (Table 13). This result, similar to that obtained in Experiment I, again indicates that course grades and aptitude scores are probably not effective predictors of team performance. As in Experiment I, the teams generally improved with practice.

EXPERIMENT III. RESOLVING TEAMMATE DISAGREEMENTS (MODULE 4)

Experimental Objectives

The primary objective of the third experiment was to determine whether a third man could improve team performance through resolving the disagreements of two teammates. For the purposes of the experiment, disagreements included those identifications unique to the initial interpreter or the checker as well as those identifications about which the other two teammates disagreed. Three resolution conditions were employed in the experiment:

Condition A. Annotated Imagery Only--Pre-Discussion. The third man interpreted all target items about which the two other team members disagreed at the end of the checking phase. The third man had available the annotated imagery but not his teammates' identifications.

Condition B. Complete Information--Pre-Discussion. The third man interpreted all target items about which the team disagreed at the end of the checking phase. However, in this condition, he was allowed to look at the identifications made by each team member as well as the annotations.

Condition C. Complete Information--Post-Discussion. The third man interpreted only those target items about which the team still disagreed after discussion. He was allowed to look at the identifications and annotations made by the team members.

Secondary objectives were to compare scoring rules for combining the third man's identifications with those of his teammates and to compare the productivity of two-man teams with three-man teams. The latter comparison was necessarily restricted in scope owing to the limited number of team structures and methods used in the experiment.

Experimental Design

A replicated 3 x 3 x 3 Latin-Square design was used. There were three resolution conditions, three missions (or periods), and three teams (or orders). The square was replicated four times, using a total of 12 three-man teams. Thirty-six officers from two image interpreter classes were used to form the teams. Unlike Experiments I and II, the subjects in Experiment III were assigned randomly to the teams in the squares.

Team Procedures

The team procedures consisted of subsets of activities or modules as in the first two experiments. Modules 1 and 2, the independent interpretation and checking modules, were identical to those employed earlier, with one exception: During the checking module, all teams used Condition B of the first experiment. That is, the checker was allowed to see only his teammate's annotations. Only two men in a team performed the first two modules; the third man worked on unrelated, non-scored imagery during this time. The third man entered the team operations prior to any discussion under two of the experimental conditions (A and B) and after the discussion module of Condition C. He devoted his attention entirely to the identifications upon which the two original teammates had failed to agree and did not look for any additional targets.

Team Scoring Rules

The checker and two-man agreement scoring rules adopted in Experiments I and II were used in this experiment to score the pre-discussion team product. Slight modifications of these rules were used to score the third man's attempts at resolution:

Third Man Final (Arbitrary). Whatever responses the third man made were scored and added to the agreed upon identifications produced in Modules 1 and 2.

Consensus (two out of three). The third man's responses was added to the agreed upon identifications produced in Modules 1 and 2 only if he agreed with either of the first two men. If he disagreed with both men, the response was thrown out.

Team Methods

The scoring rules, when applied to the team procedures employed in Experiment III, resulted in eight team methods of which the following six were used in the analysis: checker pre-discussion, two-man agreement pre-discussion, third man final pre-discussion, third man final post-discussion, third man consensus pre-discussion, and third man consensus post-discussion.

Results of Experiment III

The effect of the three resolution conditions on the performance variables may be seen in Table 20 which shows the mean accuracy, completeness, total error, and efficiency scores of the teams for both the third man final and two-out-of-three consensus scoring rules. None of the mean scores were significantly different among the resolution conditions (see analysis of variance, Tables 21 and 22).

The mean performance scores obtained for the four pre-discussion team methods are shown in Table 23. Values were obtained by summing team identifications across missions for the pre-discussion resolution conditions (A and B). (The complete information post-discussion resolution condition values were not included in this analysis.) The scores obtained for the four team methods were considered replication scores in the analysis of variance (see Table 24). The analysis essentially compared two-man performance with three-man performance and the arbitrary checker and third man final scoring rules with the consensual two-man and two-out-of-three rules. Significant differences on total error and completeness were obtained in favor of three-man teams. Two-man teams were significantly more efficient, while no difference was obtained on accuracy. The results of the analyses for the scoring rules were again identical to results from the other two experiments: The arbitrary checker rule produced significantly better performance on total error, efficiency, and completeness, while a consensus rule led to more accurate team output.

Table 20

MEAN PERFORMANCE SCORES FOR RESOLUTION CONDITIONS UNDER THIRD MAN FINAL AND CONSENSUS METHODS (Experiment III)

Resolution Condition	Accuracy	Completeness	Total Error	Efficiency
Third Man Final				
Annotated Imagery (Pre)	84%	46%	34	.21
Complete Knowledge (Pre)	90%	45%	34	.21
Complete Knowledge (Post)	87%	49%	32	.19
Consensus				
Annotated Imagery (Pre)	87%	46%	33	.21
Complete Knowledge (Pre)	93%	45%	34	.21
Complete Knowledge (Post)	88%	49%	32	.19

Table 21

SOURCE OF VARIATION, MEAN SQUARES, AND F-RATIOS FOR RESOLUTION CONDITIONS UNDER THIRD MAN FINAL TEAM METHOD
(Experiment III)

Source of Variation	d.f.	Accuracy	Completeness	Total Error	Efficiency
Orders	2	.31	1.61	.39	.81
Teams (Mean Square)	9	.014	.0039	25.09	.0013
Resolution Condition	2	.99	.02	.04	.23
Periods	2	.01	3.[illegible]	3.23	2.66
Latin Square Error	2	.26	.[illegible]	.00	.05
Resolution Procedures x Teams (Mean Square)	18	.010	.023	365.78	.012

Table 22

SOURCE OF VARIATION, MEAN SQUARES, AND F-RATIOS FOR RESOLUTION CONDITIONS (THIRD MAN CONSENSUS TEAM METHOD)
(Experiment III)

Source of Variation	d.f.	Accuracy	Completeness	Total Error	Efficiency
Orders	2	.09	1.88	.75	.91
Teams (Mean Square)	9	.012	.0031	21.36	.0014
Resolution Conditions	2	1.10	.17	.03	.24
Periods	2	.06	3.26	3.32	2.55
Latin Square Error	2	.22	.06	.02	.02
Resolution Procedure x Teams (Mean Square)	18	.0096	.024	370.08	.013

Table 23

MEAN PERFORMANCE SCORES FOR TEAM METHODS FOR RESOLUTION CONDITIONS A AND B COMBINED*
(Experiment III)

Team Method	Accuracy	Completeness	Total Error	Efficiency
Checking Pre-Discussion	85%	46%	67	.26
Two-Man Agreement Pre-Discussion	91%	40%	72	.22
Third Man Final Pre-Discussion	87%	45%	68	.21
Third Man Consensus Pre-Discussion	90%	45%	67	.21

*Two-Man vs. Three-Man and Arbitrary vs. Consensus scoring rules significantly different (P < .05) for all variable comparisons except Accuracy for the Two-Man vs. Three-Man comparison.

Table 24

SOURCE OF VARIATION, MEAN SQUARES, AND F-RATIOS FOR COMPARISON AMONG TEAM METHODS
(Experiment III)

Source of Variation	d.f.	Completeness	Accuracy	Total Error	Efficiency
Teams	11	87.9**	8.9**	156.3**	56.5**
Two-Man vs Three-Man (TT)	1	11.1**	.17	6.4*	37.7**
Arbitrary Checker vs Consensus (AC)	1	21.8**	15.8**	7.5*	16.4**
TT x AC	1	25.3**	2.6	13.0**	16.4**
Residual (Mean Square)	33	.0057	.0017	8.75	.00025

*Means significantly different, P < .05
**Means significantly different, P < .01

DISTRIBUTION

U. S. Army Personnel Research Office

DISTRIBUTION LIST

Directorate for Armed Forces I and E
Director, Army Research, OCRD
Deputy Chief of Staff for Personnel
Assistant Chief of Staff for Force Development
Assistant Chief of Staff for Intelligence
Chief of Personnel Operations, DA
CG, U. S. Continental Army Command
CG, U. S. Army Combat Development Command
CO, U. S. Army Enlisted Evaluation Center
Chief of Information, DA
Chief of Chaplains, DA
Assistant Secretary of Defense for Education
CG, Automatic Data Field Systems Command
Comdt., Marine Corps
Director, Human Resources Research Office
Directors of Research, HumRRO Field Divisions
U. S. Army Medical Research Laboratory, Psychology Division
CO and Director, U. S. Naval Training Devices Center
CG, U. S. Army CEDEC
CG, U. S. Army Electronic Proving Ground
OIC, U. S. Naval Medical NP Research Unit
Director, WRAIR, Walter Reed Army Medical Center
Chief, Personnel Research Staff, OP, U. S. Department of Agriculture
The Adjutant General's Office, Personnel Services Support Directorate
Chief of Naval Personnel
Office of Naval Research
Special Operations Research Office
Director, National Security Office
Director, Central Intelligence Agency
Chief, Office of Personnel, PHS, Department of Health, Education, and Welfare
Chief, U. S. Army R and D Office (Panama)
Office of the Provost Marshall General
Office of the Surgeon General, DA
CG, U. S. Army Materiel Command
CG, U. S. Army Security Agency
Director of Rsch and Dev., U. S. Army Electronics Command
Head, Psychology Labs., U. S. Army Natick Laboratories
CO, U. S. Army R and D Group (FE)
CG, Aberdeen Proving Ground, Hum Engr Lab
Chief, Bu M and S, Department of the Navy
Director, U. S. Naval Research Laboratory
Director, USA Engr Rsch and Dev Labs., Fort Belvoir
CO, U. S. Army Research Office (Durham)
Chief, U. S. Army R and D Liaison Group (Eur)
Chief, U. S. Army R and D Office (Alaska)

Chief, U. S. Army R and D Office, U. S. Army Arctic Test Center
CG, Air R and D Command
Chief, Officer Rsch and Review Br, U. S. Coast Guard Hq.
U. S. Army Standardization Group (Canada)
U. S. Army Standardization Group (UK)
Comdt., Command and General Staff College
Director, Military Psychology and Leadership, USMA
Superintendent, U. S. Air Force Academy
Director of Admission, U. S. Coast Guard Academy
Cmdt., U. S. Army Management School
CG, U. S. Army Infantry Center
Cmdt., U. S. Army Artillery and Missle School
Cmdt., U. S. Army Missle and Munitions Center and School
Cmdt., USAF Air Ground Operations School
Cmdt., U. S. Army Engineer School
U. S. Army Air Defense Board
U. S. Army Aviation Test Board
Director of Instruction, U. S. Army Special Warfare School
Educational Advisor, U. S. Coast Guard Training Center
Cmdt., U. S. Army War College
Cmdt., U. S. Army Aviation School
Cmdt., USASA Training Center and School
Director of Instruction, U. S. Army Armor School
Superintendent, U. S. Naval PG School
Dean, Marine Corps Institute
U. S. Army Infantry Board
U. S. Army Security Board
Library of Congress, Exchange and Gift Division
Army Library
Library of Congress, Unit X, Documents Expediting Project
Defense Documentation Center

APPENDIX

Forms Used in Experiments I, II, and III

TARGET LIST USED IN EXPERIMENTS I, II, AND III

Cargo Truck (C. Trk)
1. 1/4 Ton ($\frac{1}{4}$ T)
2. 3/4 Ton (also Ambulance) (3/4 T, AMB)
3. 2 1/2 Ton (2 $\frac{1}{2}$ T)
4. 5 Ton (5 T)
6. Dump (Dump)

Tractor Truck (Trac. Trk)
1. 5 Ton (5 T)
2. 10 Ton (10 T)

Tank Truck (Tk. Trk)
1. Water (Water)
2. Fuel (Fuel)

Wrecker Truck (Wrk. Trk)
1. 5 Ton (5 T)
2. 10 Ton (10 T)

Cargo Trailer (C. Trl)
1. 1/4 Ton ($\frac{1}{4}$ T)
2. 3/4 Ton (3/4 T)
3. 1 1/2 Ton (1 $\frac{1}{2}$ T)

Semitrailer (S/Trl)
1. Low Bed (Low Bed)
2. Tank Transporter (Tk. Tran.)
3. Tank, Gasoline (Tk. Gas.)

Support Trailer (S. Trl)
1. Water (Water)
2. Generator (Gen)
3. Ammo (Ammo)

Construction Vehicle (C. V.)
1. Grader (GDR)
2. Crane (CRN)
3. Scoop (SCP)
4. Bulldozer (BULL)

Tank (Tk)
1. M-41
2. M-48
3. M-60

Gun SP (Gun SP)
1. M-42
2. M-56
3. M-53

APC (APC)
1. M-59
2. M-113
3. M-114
4. M-75

Howitzer SP (How. SP)
1. M-108
2. M-44A1
3. M-109
4. M-55
5. M-37

Support Vehicle (S. V.)
1. Bridge Armored (BAL) Launcher
2. Recovery Vehicle (RV)

Howitzer (How)
1. M-101 A1
2. M-1 A1
3. M-2 A1

Tent (Tent)
1. General Purpose (GP)
2. Wall (Wall)
3. Pup (Pup)
4. Command Post (CP)
5. Hexagonal (Hex)
6. Maintenance (Main)
7. Kitchen (Kit)
8. Canvas Shelter (CS)

MISSION INFORMATION FORM USED IN EXPERIMENT I

INITIAL INTERPRETER				CHECKER		
TIME STARTED ______				TIME STARTED ______		
TIME FINISHED ______				TIME FINISHED ______		
NAME ______				NAME ______		
FRAME #	# OF ANNOTATIONS	# OF TARGETS		FRAME #	# OF ANNOTATIONS	# OF TARGETS

MISSION #______ SCALE ______ CONDITION ______

FRAMES ______ DATE ______

TIME DISCUSSION STARTED ______ TIME DISCUSSION FINISHED ______

INITIAL INTERPRETATION REPORT FORM USED IN EXPERIMENT I

FRAME # ______________ II INITIALS __________

1	2	3	4	
ANNOT. #	IDENTIFICATION	GRID SQUARES		
		PRIMARY	NEAREST	

CHECKING REPORT FORM USED IN EXPERIMENT I

FRAME # ______________ II INITIALS ____________

1	2	3	4	5	6	7
ANNOT. #	IDENTIFICATION	GRID SQUARES		TEAM-MATES ANNOT. #	DIS-AGREE-MENT	AGREED ID

MISSION INFORMATION FORM USED IN EXPERIMENT II

INITIAL INTERPRETER

TIME STARTED ______________

TIME FINISHED ______________

NAME ____________________

FRAME #	# OF ANNOTATIONS	# OF TARGETS

CHECKER

TIME STARTED ______________

TIME FINISHED ______________

NAME ____________________

FRAME #	# OF ANNOTATIONS	# OF TARGETS

MISSION # __________ SCALE ______________ CONDITION __________

FRAMES ________________________________ DATE __________

TIME DISCUSSION STARTED ____X____ TIME DISCUSSION FINISHED ___X___

IMAGE INTERPRETER REPORT FORM USED IN EXPERIMENT II

FRAME #__________

CHECK ALL IDENTIFICATIONS AND FRAMES WITH CONFIDENCE OF ___% OR LESS

1	2	3	4	5	6	7	8
ANNOT. #	CHECK THIS ITEM	GROSS IDENTIFICATION	CONFI-DENCE	DETAILED IDENTIFICATION	CONFI-DENCE	GRID SQUARES PRIMARY	NEAREST
		(1)					
		(2)					
		(1)					
		(2)					
		(1)					
		(2)					
		(1)					
		(2)					
		(1)					
		(2)					
		(1)					
		(2)					
		(1)					
		(2)					
		(1)					
		(2)					

CONFIDENCE THAT ALL TARGETS HAVE BEEN DETECTED ___% (II 1) ___% (II 2)

II INITIALS ______ (II 1) ______ (II 2)

MISSION INFORMATION FORM USED IN EXPERIMENT III

INITIAL INTERPRETER	CHECKER
TIME STARTED ________	TIME STARTED ________
TIME FINISHED ________	TIME FINISHED ________
NAME ________	NAME ________

DISAGREEMENT TARGETS				DISAGREEMENT TARGETS			
1	2	3	4	1	2	3	4
Frame #	Annot. #	Identification	Conf.	Frame #	Annot. #	Identification	Conf.

MISSION # ________ FRAMES ________

SCALE ________ CONDITION ________ DATE ________

FIRST DISCUSSION START TIME ________ FIRST DISCUSSION FINISH TIME ________

SECOND CHECKER START TIME ________ SECOND CHECKER FINISH TIME ________

NAME (SECOND CHECKER) ________

INITIAL INTERPRETATION REPORT FORM USED IN EXPERIMENT III

FRAME NR. ______________ II INITIALS ______________

1	2	3	4	5	
ANNOT. NR.	IDENTIFICATION	CONFIDENCE	GRID SQUARES		
			PRIMARY	NEAREST	

FT 4538h 65:4538h

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) System Development Corporation, Santa Monica, Calif. U. S. Army Personnel Research Office, Wash., D. C.	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

The Use of Teams in Image Interpretation: Information Exchange, Confidence, and Resolving Disagreements

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (Last name, first name, initial)

Doten, George W. and Cockrell, John T. (SDC) and Sadacca, Robert (U. S. APRO)

6. REPORT DATE	7a. TOTAL NO OF PAGES	7b. NO OF REFS
October 1966	57	3

8a. CONTRACT OR GRANT NO. DA-49-092-ARO-65 b. PROJECT NO. DA R&D Proj. No. 2J620901A721 c. Component Integration Task d.	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Research Report 1151 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. AVAILABILITY/LIMITATION NOTICES

Qualified requestors may obtain copies of this report directly from DDC. Available, for sale to the public, from the Clearinghouse for Federal Scientific and Technical Information, Department of Commerce, Springfield, Virginia 22151.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Assistant Chief of Staff for Intelligence Office, Chief of Research and Development

13. ABSTRACT Among the specific objectives of the COMPONENT INTEGRATION Task, U. S. Army Personnel Research Office is the identification of effective team procedures under various system conditions and requirements. In prior studies, research has focused on the basic question of whether teams can perform image interpretation more effectively than can individuals acting alone, and on related questions concerning best team methods and procedures and best size of teams for maximal performance. The present study, conducted jointly by personnel of the Advanced Systems Division, System Development Corporation, and U. S. APRO, centers on system team interactions designed to reduce the time required for team interpretation while maintaining the superiority of team procedures in the accuracy and completeness of the imagery information extracted. Three experiments were conducted, using the common procedure of having each team member in two-man teams check the interpretation of his teammate. Three specific primary objectives were established: 1) To determine the amount and type of knowledge which the checker should have of the initial interpreter's work; 2) To determine whether the initial interpreter can accurately judge when his work needs to be checked by his teammate; and 3) To determine how best to utilize a third man to resolve disagreements among teammates on interpreted items. Variations in procedures were achieved for analysis by setting up four phases or modules of interpreter team activity. Team results produced under each method were assessed in terms of completeness, amount of error, accuracy, and efficiency. Findings suggest that 1) more complete results are produced with higher efficiency in teams where the checker has full knowledge of the initial interpreter's work; 2) only limited judgment as to the adequacy of their interpretations can be made by initial interpreters;

DD FORM 1 JAN 64 1473

Unclassified

Security Classification

Unclassified
Security Classification

14 KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
Image systems--work methods						
*Information extraction						
*Surveillance displays						
*Aerial surveillance						
*Image interpreter team operation						
*Team methods						
Laboratory facilities						
Image interpreter performance measures						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
______________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
______________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
______________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

Unclassified
Security Classification

DD Form 1473

13. ABSTRACT continued

3) team performance increases in completeness but decreases in efficiency with the introduction of a third man; 4) results with different team methods pose a tradeoff situation, since no one method appears to hold best for team performance under all requirements.

B15323