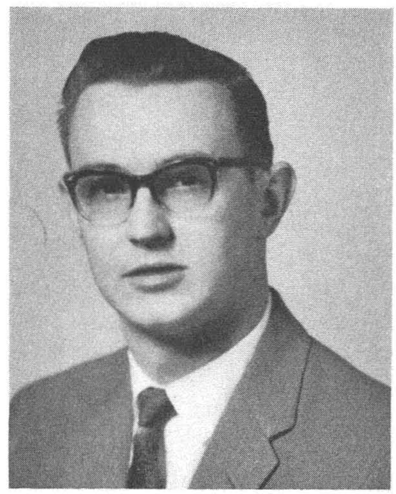
*Human Factors Studies in Image Interpretation**

Part I
Introduction†

HAROLD MARTINEK,
and
ROBERT SADACCA
Human Factors Research Branch,
The Adjutant General's Research and Development Command,
U. S. Army

IN ORDER to determine the human factors requirements in image interpretation systems, the Human Factors Research Branch of The Adjutant General's Research and Development Command, U. S. Army conducted an extensive preliminary survey to define image interpretation problems which would clearly focus on the Army's needs and which, at the same time, would be amenable to research attack. A statement of the military requirements in image interpretation and the formulation of HFRB's research program in this area were a direct outgrowth of the survey.



HAROLD MARTINEK

From its photo interpreters the Army needs information which is timely, relevant, complete, and accurate. Until recently, the Army's photo interpreter was confronted only with the problem of interpreting relatively large-scale, conventional black-and-white photographs. Today he is also required to interpret small-scale and degraded photographs and radar, infra-red, and TV imagery. All the military services have developed—and are still developing—their capacity to take aerial photographs from a variety of aerial platforms. They have also increased tremendously their capacity to process imagery at a rapid rate. This increased capacity and variety of image sources have inevitably increased the quality range of images. But good or poor, the imagery obtained through these media is ultimately placed before the photo interpreter who is asked to extract information useful for making important military decisions.

Existing knowledge of the basic psychological factors operating when photo interpreters interpret real imagery is severely limited. HFRB's research program is being formulated to answer two broad questions: (1) What are the skills, abilities and techniques necessary to extract intelligence in-

* The four papers in this Symposium were presented at the Society's 27th Annual Meeting, The Shoreham Hotel, Washington, D. C., March 19–22, 1961.

† The opinions expressed in this paper are those of the authors and do not necessarily reflect official Department of the Army policy.

AUTHOR'S NOTE for all four papers in the Symposium. Appreciation is expressed to Lt. Col. Matthew J. Toia and to Lt. Col. Eugene H. Manley of the U. S. Army Intelligence Center for assistance in the formulation of the problems and development of test materials. Appreciation is also expressed to the Image Interpretation Section, 525th MI Group (Lt. Robert Perry, Section OIC) for assistance in image selection, historical research, and development of performance measures.

formation from conventional and newer types of imagery? (2) How can the Army best utilize its available human resources in order to cope with the ever-increasing variety of image types, and at the same time maintain acceptable standards of speed, accuracy, and completeness? Most other research efforts have concentrated upon improving the nature of the stimuli presented to the photo interpreter, that is, on improving the quality of obtained imagery. In HFRB's formulation of the Army's human factors image interpretation problem, there has been accepted the fact that poor quality imagery will at times be the only imagery available and may urgently require interpretation. It is hoped to complement research directed toward improving the quality of imagery by conducting research directed toward selecting better photo interpreters and improving the procedures and techniques they employ, the composition of teams and team-work methods, and the way in which they use their equipment.

In order to conduct the many studies needed to answer the formulated research questions, the initial efforts have been devoted primarily toward the development and tryout of a series of performance measures. These measures will be the instruments or vehicles by means of which analysis will be made of the psychological dimensions of Army Photo interpretation. In these measures attempt will be made to reproduce as far as possible typical real-life situations which confront the photo interpreter. Thus the imagery selected for inclusion in the measures is from actual tactical and strategic air reconnaissance missions and, in the case of photography taken during war time, frequently shows enemy-held positions. The objects the photo interpreter is directed to look for reflect the intelligence requirements of the military situation at the time of the air reconnaissance mission. As in an operational situation, the interpreter is provided with maps of the relevant areas, sortie plot overlays, and general and specific information about the location and disposition of friendly and enemy forces. Also, as in the operational situation, the photo interpreter is allowed to examine the imagery presented to him in any way he desires, and is allowed to use any of the instruments in his PI kit while making his interpretations.

While these measures are being developed and tested small pilot studies into selected human factors aspects of photo interpretation are being conducted. For one thing, these pilot studies permit noting the reactions of PI's to the performance measures, and to iron out problems in administering, scoring, and analyzing the measures. More important, the pilot studies assist in determining the further research value of some of the HFRB hypotheses and the priorities to be given various research projects. Also, since these pilot studies are directed at critical problems relating to ground data processing of intelligence information, the results of these studies may have immediate application in the military setting. In addition to the three pilot studies reported below, HFRB is currently investigating: (1) the development of a selection battery, (2) optimal group composition and work procedures of tactical PI units, and (3) reliability of the interpretation report.

Now, concerning the material used and how it was used in the studies which follow and indeed most of the other studies which have been run, certain communalities exist. These are: (1) the photography used, (2) the determination of ground truth, and (3) the experimental controls.

The photography used in the experiments was generally quite good. Only original negatives were used and the print quality was such that all subjects agreed that the quality was very good to excellent and in general better than they were accustomed to working with. The scale varied from 1:4,000 to 1:8,000 for tactical interpretation problems and from 1:10,000 to 1:15,000 for strategic interpretation problems. These scales at the very least are sufficient for interpretation of the objects on the ground.

In determining the content of the photographs, or the ground truth, different approaches were used on the tactical and strategic photography. For the tactical, historical records of the Korean Action were studied and related to the photography. Using the knowledge obtained from these records several experienced photo interpreters of the 525th MI Group interpreted the photographs. The interpretations were checked by two experienced interpreters of HFRB. A third check was made by instructors at the USAINTS. The few objects which the above three groups could not agree on were scored as neither right nor wrong. In a few instances of the tactical photography, no background information was available, and the ground truth was determined solely by the triple interpretations of experienced photo interpreters.

The ground truth of the strategic photography was perhaps more easily determined from maps, the bombing encyclopedia, and other sources by three experienced strategic interpreters of HFRB. These were checked as before by instructors of the USAINTS. Any object or installation actually existing but not clearly recognizable by the experts was scored as neither right nor wrong.

In the three studies which follow, comparisons were made in each case between or among matched groups of interpreters. The matching of the groups was accomplished by using two matching variables: their final course grades and their score on several predictor tests administered before they started training. Thus, it is felt that the groups were equated on PI content learned at school and general ability level. In addition, except for a couple of subjects in one of the studies, the ratio of officers to enlisted men was constant across groups in any one study. Other controls used were: (1) each group did not know what the other groups were doing or that their work was to be compared to the other groups, (2) each subject reported only those objects indicated on standard object lists, using only the exact words given, and (3) statistical tests of significance of differences were computed where possible.

As will become apparent, the studies did not reflect all operational conditions typical of actual interpretation. Of course, the general feeling of one associated with war time conditions could not be artificially reproduced. Other limitations of these pilot studies include, (1) the lack of comparative cover, (2) the lack of keys, (3) the lack of use of the complete range of scale and quality of photography, (4) the lack of all objects and situations found in war time, and (5) the lack of data on experienced photo interpreters. These limitations however do not bias our results but only restrict their generality. More studies obviously have to be accomplished before a firm decision can be made concerning the adoption of the findings.

In summary, the pilot studies which follow are related to only part of the total HFRB program of research in the area of image interpretation. In general, the research program is designed to empirically optimize such human factors areas as selection methods, PI administrative and technical procedures and techniques, team approaches, and use of tools and keys in order to maximize the speed, accuracy, and completeness of intelligence information derived from the various aerial sensors.

Part II
Comparison of Photo Interpreter Performance in Separate and Combined Viewing of Vertical and Oblique Photography†

JOHN E. RANES,
Human Factors Research Branch,
The Adjutant General's Research and Development Command,
U. S. Army

Introduction

In the identification of the psychological processes and procedures utilized in image interpretation, the various image inputs to the interpreter must be considered in terms of image type and image combinations. That combination most familiar to the interpreter today is the one providing him with more than one view of the target area, the vertical and oblique cover combination. In the operational situation there frequently has been available to the photo interpreter both vertical and oblique cover of the same tactical or strategic situation. Experienced photo inter-

† The opinions expressed in this paper are those of the author and do not necessarily reflect official Department of the Army policy.

JOHN E. RANES

preters indicate that the location and identification of objects is often aided by both types of cover and that objects suggested by one type of photography can be confirmed or denied with that taken from a different view-angle.

BACKGROUND

In considering this particular image input to the interpreter, several questions were posed in this study. For example:

a. Are two views of the target area better than one?
b. When two views are available, is the better procedure that in which the interpreter looks freely at either view of the target area as he wishes, or that in which he looks at first one type of cover, and then confirms his interpretations with the other type of cover?
c. Is it better procedure for the interpreter to first look at oblique cover, then later the vertical cover to check all of his original interpretations; or to first look at the vertical cover and then the oblique?
d. If only a single view is available, are there any performance differences between vertical and oblique viewing?

In order to gain some insight into these problems, an experimental pilot study was conducted at the U. S. Army Intelligence School from July through November 1960, using as subjects 109 graduates from the photo interpretation courses (36 officers and 73 enlisted men). These graduates were divided into five matched groups on the basis of their final grade averages and a composite of their aptitude test scores, with officers and enlisted men divided about evenly among the groups. The vertical and oblique photos selected for this experiment were obtained from trimetrogon cover of a North Korean convoy early in the Korean War. The convoy appeared in both the vertical cover and in the left oblique cover. Any one group received either one type of imagery, both types simultaneously, or both types in different sequence.

Groups 1 and 5 of Table 1 were given the vertical cover and groups 2 and 4 the oblique views. Group 3 received both the vertical and oblique cover. All groups examined the imagery for 15 minutes and completed reports on the vehicles in the convoy. Then groups 3, 4, and 5 were told to continue examining their imagery for 15 more minutes and to make any additions or deletions they felt were appropriate. Groups 1 and 2 were given the same instructions; however they were given the additional view of the target area, the oblique cover in the case of group 1 and the vertical cover in the case of group 2.

TEST ADMINISTRATION

The subjects were placed in the proper setting by providing them with a situation sheet, the appropriate photography for their group, a sortie plot overlay, and the appropriate map cover. The situation sheet in this vertical and oblique cover experiment gave information to the subject that his team chief had scanned photography received that day and had selected specific photos for further interpretation. Specific requirements were to locate and identify all vehicles on or near the road within brackets annotated on the photography by his team chief. On being provided the cover for the first 15 minute period none of the groups were informed that an additional 15 minutes would be allowed. Those groups provided with only one view (oblique or vertical) were not informed that another viewing angle would be provided.

TABLE 1

VERTICAL (V) AND OBLIQUE (O) IMAGERY AVAILABLE DURING TWO TIME PERIODS FOR EACH GROUP

	1	*2*	*3*	*4*	*5*
First 15 minutes	V	O	V & O	O	V
Second 15 minutes	V & O	V & O	V & O	O	V

In taking the tests, the photo interpreters annotated by number directly on the photographs the objects (in this case—vehicles) they had located, and then recorded their identifications on special answer sheets using only the descriptive terminology allowed. In this experiment a general list of vehicles, both tracked and wheeled, was provided with the situation sheet, and only gross identification was required such as "truck," "car," "tank," etc. The subjects also recorded the confidence they felt on each identification, using a three-point scale (positive, probable, and possible). At the end of every 5 minute time period the subjects marked their answer sheet beside the last entry made in order to provide information for a time analysis. Special instructions provided in all test administrations requested that no identification be changed or erased, however, the subject could use the next line on his answer sheet to change or omit by number a previously recorded item.

Analysis of Data

Various statistical analyses were made to test for the significance of differences in means among the five groups. Performance scores obtained for evaluation of the various group differences were *right* scores, *wrong* scores and *accuracy* scores, with *accuracy* defined as the number of *right* identifications divided by the sum of *right* and *wrong* identifications. Thus, for each of the time periods (the first 15 minutes, the second 15 minutes, and the total 30 minute period) three scores were analyzed by statistical tests of significance. Of fourteen (14) identifiable vehicles in the imagery, none defied detection. Each vehicle was reported by at least one of the subjects examined, and at least one subject reported all fourteen (14) vehicles.

As shown in Table 2 the average number of correct identifications among the five groups in the first 15 minute period was 7.5 (ranging from 6.5 to 8.4). The similar average for the second 15 minutes was .9 (ranging from .7 to 1.1). For the total 30 minute period the average number of correct identifications was 8.4 (ranging from 7.6 to 9.4).

The analyses of *right* scores revealed *no* statistically significant differences, among the group means for all time periods, either in the type of imagery viewed or in the order of presentation; thus, the slight variations in the group mean scores are considered to have occurred by chance. Thus, there were no differences in performance between vertical viewing alone as opposed to simultaneous viewing of vertical and oblique.

The analysis of incorrect identifications, however, did indicate differences in performance. *Wrong* scores for the first 15 minute period did not reveal any statistically significant differences, with the average number of *wrong* identifications among the five groups being 6.5 (ranging from 5.1 to 7.5). (See Table 3.) However, during the second 15 minute period the group comparisons revealed that those groups receiving additional imagery (groups 1 and 2) produced almost twice the number of *wrong* identifications as did the other groups, There was a statistically significant difference in the comparison between groups receiving or *not* receiving additional imagery (comparison of groups 1 and 2 with groups 3, 4, and 5), but not in the comparison of order of presentation of the oblique and vertical imagery (comparison of group 1 with group 2).

Table 3 shows that for the entire 30 minute period, the average number of *wrong* identifications among the five groups was 8.3, varying from 6.3 in group 4 (oblique imagery only) to 9.5 in group 2 (which received oblique in the first period and vertical in the second period). The comparison of group 4 (oblique imagery only) with the other four groups (which had all received vertical imagery for a part of the total period at least) also revealed a statistically significant difference.

Table 2

Comparisons of Mean Right Scores in Vertical and Oblique Viewing

Group	*First 15 Minutes*		*Second 15 Minutes*		*Total 30 Minutes*
1	(V)	7.9	(V+O)	.9	8.8
2	(O)	6.5	(O+V)	1.1	7.6
3	(VO)	8.4	(VO)	1.0	9.4
4	(O)	7.3	(O)	.8	8.1
5	(V)	7.3	(V)	.7	8.0
Average of Groups:		7.5		.9	8.4

TABLE 3

COMPARISONS OF MEAN *Wrong* SCORES IN VERTICAL AND OBLIQUE VIEWING

Group	*First 15 Minutes*		*Second 15 Minutes*		*Total 30 Minutes*
1	(V)	6.5	(V+O)	2.0	8.5
2	(O)	6.7	(O+V)	2.8	9.5
3	(VO)	7.5	(VO)	1.6	9.1
4	(O)	5.1	(O)	1.2	6.3
5	(V)	6.7	(V)	1.3	8.0
Average of Groups:		6.5		1.7	8.3

As shown by Table 4 the accuracy comparisons revealed no statistically significant differences in any of the three time periods. Among the five groups the average, for the first 15 minute period was 54%, ranging from 49% to 59%. The high number of *wrong* scores versus the lower number of *right* scores in the second 15 minute period resulted in a considerable drop in accuracy, with an average among the five groups for this period of 34%, with a range of 28% to 40%. For the total 30 minute period the average accuracy was 50%, varying from 44% to 56%. Although the largest drop in accuracy occurred in the two groups that had received the additional imagery during the second 15 minute period, it was not significantly different.

SUMMARY

Before summarizing these results, it must first be emphasized that they are of a tentative nature. They are based on samples of inexperienced photo interpreters and only one experimental performance measure, limited in content to vehicles in convoy. The data obtained thus far in this pilot study, however, do point to some interesting differences that may generalize to other situations in viewing and interpreting from imagery obtained from different view points.

In summary, 109 school trained but inexperienced photo interpreters were divided into five matched groups. For the entire 30 minute test period one group received vertical photos, a second group received oblique photos, and a third group received both vertical and oblique photos. The remaining two groups received either vertical or oblique photos for the first 15 minutes and both views for the second 15 minutes.

Statistically significant differences were found in two instances in the analyses of *right*, *wrong*, and *accuracy* scores. The first difference was observed in the comparison of *wrong* scores from groups interpreting from the same packet of imagery for the entire 30 minute period and groups receiving additional imagery in the second half of the period. The second difference occurred in the comparison or *wrong* scores of groups receiving oblique imagery alone and groups receiving the vertical view for at least half of the period.

Tentative results of this pilot study indicate that:

a. Higher *wrong* scores resulted when additional imagery was presented halfway through the time period.

b. The group using oblique imagery alone produced lower *wrong* scores for the total 30 minute period than did the other groups.

c. Simultaneous use of both views of the target area compared with vertical alone viewing yielded no statistically significant differences.

TABLE 4

COMPARISONS OF MEAN *Accuracy* IN VERTICAL AND OBLIQUE VIEWING

Group	*First 15 Minutes*		*Second 15 Minutes*		*Total 30 Minutes*
1	(V)	.55	(V+O)	.31	.51
2	(O)	.49	(O+V)	.28	.44
3	(VO)	.53	(VO)	.38	.50
4	(O)	.59	(O)	.40	.56
5	(V)	.52	(V)	.35	.50
Average of Groups:		.54		.34	.50

d. Changing the order of presentation of the imagery, vertical first or oblique first, yielded no significant differences.

If these results on vehicle identifications hold for other types of objects and photo qualities, they would indicate to the photo interpreter that he does not necessarily need to obtain the additional view of the target area to confirm or deny his identifications. More extensive study could provide to the interpreter *and* to the air reconnaissance planning personnel specific guidance in the collection of imagery of particular types of target areas consistent with intelligence needs. The techniques followed in the conduct of this pilot study may well be applicable to studies of photo interpreter performance in other image combinations obtained from airborne sensors, to include the returns from side-looking radar, infra-red, television, and the conventional aerial photograph.

To continue with this particular pilot study, additional performance measures will be developed using vertical and oblique cover of objects in several categories, such as weapons, other types of vehicles, structures and facilities, and other man-made objects. This cover will also be obtained at different scales. These performance measures will be administered to graduating photo interpreter students and to experienced photo interpreters in field organizations in order to provide more conclusive results.

Part III
Comparison of Photo Interpreter Performance Under Stereo and Non-Stereo Viewing Conditions†

ALFRED I. SCHWARTZ *and* JOSEPH ZEIDNER,
Human Factors Research Branch,
The Adjutant General's Research and Development Command

INTRODUCTION

In support of advanced reconnaissance systems, a large number of development and cost problems are incurred in order to secure stereo photographs. These problems are concerned with the determination of effective and economical means of obtaining, processing, annotating, printing and storing stereo photographs in data processing systems. Even within conventional photo reconnaissance systems, a substantial portion of over-all cost relates directly to the problem of using stereo photographs over and above simple coverage. In those situations where it can be demonstrated that stereo viewing contributes significantly to photo interpreter performance, then the expense for the development and use of the stereo capability is of course justified. In addition, the requirement for stereo viewing in the design specifications for radar and infra-red sensors is an important ongoing problem that needs to be resolved.

Findings of a previous pilot study* seemed to raise a question of the value of stereo viewing in the interpretation of small-scale strategic photographs. These findings provided the impetus for further research to assess the value of stereo viewing within different intelligence environments. While many types of stereo studies have been conducted, none provided findings that could be generalized to real-life Army photo interpreter systems.

The primary objective of the present study is to determine whether or not stereo viewing helps in the identification of militarily sig-

* "Preliminary Study Concerning the Value of Stereo Viewing for Photo Interpreter Performance," Broadview Research Corporation, Washington, D. C., March 1960.

† The opinions expressed in this paper are those of the authors and do not necessarily reflect official Department of the Army policy.

ALFRED I. SCHWARTZ

nificant tactical and strategic objects and situations that appear in photographs available in an operational setting. No attempt was made in this study to investigate systematically such variables as scale, resolution, and content. The present study was planned primarily to determine the need and direction for more refined research in this area.

In July and August 1960, data were collected on 59 individuals—21 officers and 38 enlisted men—about to graduate from the Photo Interpreter Course at the Army Intelligence Center. Two matched groups were established on the basis of aptitude tests and final course grades. Three sets of tactical photographs obtained during the Korean War and one set of strategic photographs were presented to all interpreters under simulated operational conditions. For each set of photographs, stereo pairs were provided to one of the two matched groups, and non-stereo photographs to the other. Both groups of interpreters served as either the experimental or control group for two sets of photographs. The time limits for the different measures were considered reasonable in view of the nature of the photographs and the items the PI's were asked to locate and identify.

In addition to photographs, maps, overlays, and standard references, the interpreters were provided with a situation sheet which was read aloud to them before they began to look at the photographs. The interpreters were asked to detect and identify objects of military significance. The objects of concern were provided on a list to insure standard scoreable responses to a desired level of detail. The objects were similar to those requiring identification in a real combat setting and include such targets as tanks, artillery, and fortifications.

The subjects' *right* scores consisted of the number of keyed objects correctly located and identified. Their *wrong* scores were made up of the number of objects misidentified. Most of the subjects' wrong scores did not result from misidentifying the objects in the key, but rather from the subjects identifying as valid items, objects which really had no military significance.

QUALITY OF INFORMATION

Table 1 presents the mean stereo and non-stereo responses of all individuals for each of the four measures, for the total time period allowed. Neither the mean differences for right or wrong scores were significantly different at the .05 level when the tests were made. Furthermore, there was no consistent pattern or trend favoring either method. Stereo responses were "better" four times; non-stereo, three times, and there was one tie. Actually the mean "Right" scores for the two methods were very close to one another—the largest difference being .8 (for Measure 2) in favor of non-stereo.

This same analysis was also run for an initial time period simulating flash or immediate type reporting. Table 2 presents these findings. About $\frac{1}{3}$ of the total permitted time for each measure was allowed for this type of reporting. The results for this analysis did show a consistent trend in favor of non-stereo viewing for the *right* responses, and a consistent trend in favor of stereo viewing for the *wrong* responses. However, only one of the eight comparisons was statistically significant at the .05 level. Thus there is an indication that

TABLE 1

COMPARISON OF MEAN STEREO AND NON-STEREO RESPONSES

Total Time

Measure	*Method*	*Right*	*Wrong*
1	Stereo	2.3	15.6
	Non-Stereo	2.3	16.7
2	Stereo	4.0	9.5
	Non-Stereo	4.8	8.7
3	Stereo	5.6	14.0
	Non-Stereo	5.4	15.7
4	Stereo	5.4	2.9
	Non-Stereo	5.8	4.0

information of the type required in this study perhaps can be more rapidly extracted within relatively short time periods without the use of a stereoscope, and with no discernible loss of over-all accuracy of interpretation. (A later section of this paper will deal more fully with the problem of rate of information extraction.)

An additional analysis was run by individual target type within each set of photographs. This analysis was made to determine if stereo/non-stereo differences might have been obscured by averaging all responses together. Tests of significance between stereo and non-stereo median differences for each target type within each performance test were made, both for the initial and total time periods, and for both right and wrong responses.

It was found that for the 138 comparisons made, only 14 were significant at the .05 level. Nine of these comparisons favored stereo and five favored non-stereo. By and large no discernible advantage could be detected for either of the two methods for any target type. However, there was an indication that interpreters using stereo tended to make slightly fewer errors by mistakenly calling various structures or natural features vehicles. It was further determined that of all the target types identified in the photos, none defied identification without stereo.

TABLE 2

COMPARISON OF MEAN STEREO AND NON-STEREO RESPONSES

Initial Time

Measure	*Method*	*Right*	*Wrong*
1	Stereo	1.1	5.3
	Non-Stereo	1.3	6.4
2	Stereo	1.7	2.7
	Non-Stereo	2.6	3.9
3	Stereo	3.3	5.2
	Non-Stereo	3.9	7.9*
4	Stereo	3.1	1.1
	Non-Stereo	4.0	1.4

* Significant at .05 level.

RATE OF INFORMATION EXTRACTION

One of the advantages often attributed to stereo viewing is that it enables interpreters to extract information more rapidly. In Figures 1 and 2 are shown data relevant to the rate of information extraction as a function of viewing method. Figure 1 represents the cumulative right, wrong, and total response functions over time for both stereo and non-stereo viewing conditions. Figure 2 shows the cumulative accuracy function for both stereo and non-stereo viewing. The functions shown are for one of the performance situations considered representative of the other three.

Figure 1 shows that the cumulative total number of responses made is greater under the non-stereo condition than under the stereo condition. (This is the case for the other performance tests as well.) It is also interesting to note that after 30 minutes, responses under both viewing conditions are still being made, although at a somewhat diminishing rate. From examining the wrong and right response curves, it can be seen that after about 15 minutes an increasingly greater proportion of all responses made are wrongs. This can be seen more clearly by examining the accuracy curves presented in Figure 2. Here non-stereo accuracy (defined as $R/[R+W]$) drops from about 48% to 38% in the last 15 minutes, and from 32% to 31% accuracy for stereo. It is especially noteworthy that for this performance test, non-stereo viewing consistently provides a better over-all accuracy than does stereo viewing. With several exceptions, this is true for all measures and for all time periods. However, the differences in accuracy levels for the two viewing methods generally tend to favor non-stereo viewing more during the earlier than over the later time periods.

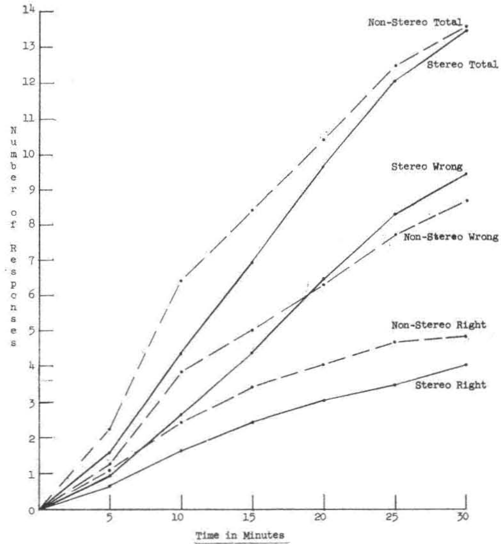


FIG. 1. Cumulative numbers of right, wrong, and tota responses by five minute intervals for stereo and non-stereo viewing conditions for measure 2.

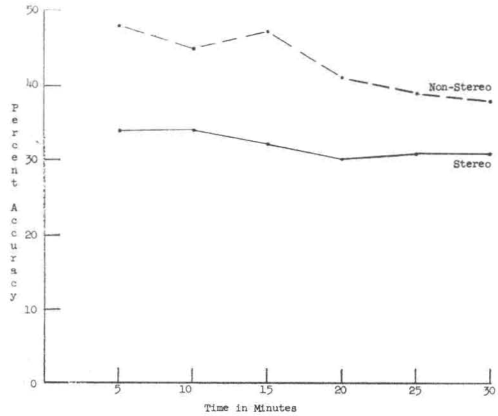


FIG. 2. Per cent accuracy at consecutive five minute intervals for stereo and non-stereo viewing conditions for measure 2.

Considering all of these time functions for all targets, it appears that unlimited time does not work to the advantage of the interpreter; right responses seem to drop off much more rapidly than wrong responses. For the targets employed under the conditions of this study, the better accuracy scores were produced in the first 10 to 15 minutes. The data also seem to indicate that for shorter time periods, accuracy might be greater under non-stereo viewing conditions.

LEVEL OF CONFIDENCE

An additional advantage attributed to stereo viewing is believed to be that it enables the interpreter to have greater confidence in his identifications. In the operational situation an interpreter normally indicates his level of certitude in his report by indicating the following qualifiers: "Positive," "Probable," and "Possible." Interpreters feel that stereo viewing enables them to give more "Positives" or fewer "Possibles" with their identifications. An additional objective of this study was to determine if any differences in confidence levels occurred as a function of the method of viewing, and also to determine if there were differences in confidence associated with right and wrong responses for each of the viewing methods.

In this study, a confidence value was assigned by each interpreter to each response he made. A value of 1 was assigned for a response called positive (greatest confidence), 2 for a probable, and 3 for possible (least confidence).

In Table 3 the mean confidence levels are shown for each measure for both stereo and non-stereo viewing conditions, broken down into right and wrong responses. It is immediately apparent that the mean differences in confidence level by viewing methods are very close in all cases. An over-all mean confidence level was obtained for each of the two viewing methods, and the difference tested for statistical significance. The difference was found not to be significant.

TABLE 3

COMPARISON OF MEAN CONFIDENCE VALUES ASSIGNED TO STEREO AND NON-STEREO RESPONSES

Measure	*Method*	*Right*	*Wrong*
1	Stereo	1.5	1.7
	Non-Stereo	1.4	2.1
2	Stereo	1.3	2.0
	Non-Stereo	1.3	2.0
3	Stereo	1.5	1.8
	Non-Stereo	1.6	2.0
4	Stereo	1.3	1.8
	Non-Stereo	1.3	1.6

Upon examination of Table 3, it can be seen that there is a consistent difference in mean confidence for right and wrong responses for both viewing methods. Fortunately the confidence expressed for responses that later were found to be right was greater than the confidence expressed for wrong responses. An over-all mean confidence level was obtained for right and wrong responses across viewing methods and measures, and the difference tested for statistical significance. On testing, this mean difference was found to be statistically significant.

The findings with regard to confidence levels indicate that interpreters actually had equal confidence in their identifications regardless of whether or not a stereoscope was employed. However, they exhibited less confidence in those responses which later turned out to be wrong than they did for right responses, regardless of viewing method. The magnitude of the differences in expressed confidence level for right and wrong responses warrants attention within the operational setting.

SUMMARY AND CONCLUSIONS

The major objective of this study was to determine whether or not stereo viewing helps in the identification of militarily significant targets that appear in photographs available in an operational setting. Data were collected on 59 officers and enlisted men about to graduate from the Photo Interpreter Course at the Army Intelligence Center. Two matched groups of interpreters were established on the basis of aptitude tests and final course grades. Three sets of tactical photographs obtained during the Korean War and one set of strategic photographs were presented to all interpreters under simulated operational conditions. For each set of photographs, stereo pairs were provided to one of the two matched groups, and non-stereo photographs to the other. The time limits for the different measures were considered reasonable.

The major findings were:

1. No significant mean differences between stereo and non-stereo viewing were found for either right or wrong scores.
2. There was no consistent pattern or trend favoring either stereo or non-stereo viewing for total score.
3. For an initial time period simulating flash reporting, there was a consistent trend in favor of non-stereo viewing for right responses, and in favor of stereo viewing for wrong responses. However, only one of these differences was statistically significant.
4. With one possible exception, no discernible advantage was detected for either of the two viewing methods for any specific target or object type.
5. Within the time limits used in this study, the total number of responses made after any given period of time is greater under the non-stereo condition than under the stereo condition, and with few exceptions without loss in overall accuracy. These differences were tested for significance at the terminal time period only at which time the differences were not found to be statistically significant.
6. The confidence level assigned by interpreters in their responses is similar for the two viewing methods. However, interpreters had less confidence in those identifications which later turned out to be wrong than they had for right responses, regardless of viewing method.

All of the findings presented concerning the value of stereo viewing must be evaluated in the light of the study setting. There are many factors that tend to limit the generalizability of the findings including level of experience of examinees, the nature of the questions posed to the interpreter, and the specificity of photo-quality and content. Nevertheless, it is concluded on the basis of these unusual findings that the contribution of stereo to location and identification of targets should be studied intensively in a new series of experiments. In these new studies, the sample should include experienced as well as newly trained interpreters; and photo content and quality variables including conditions of camouflage and concealment should be systematically controlled. The role of stereo viewing in counting and mensuration should also be examined. Through such studies, it may be possible to determine the specific kinds of interpreter problems and photo qualities for which stereo viewing could be profitably employed. If such selected application of stereo could be made, this would result in an enormous saving of time, effort, and money in all aspects of procuring, processing, and interpreting aerial photographs; further, it would provide guidance in establishing design specifications for advanced reconnaissance systems.

Part IV
The Effect of Auxiliary Intelligence Information on PI Performance†

ANTHONY E. CASTELNOVO
Human Factors Research Branch,
The Adjutant General's Research and Development Command,
U. S. Army

THE schema used herein for studying the effect of auxiliary intelligence information in PI performance is adapted from Solley and Murphy (1960) and may be diagrammed as shown in Figure 1.

The psychological dynamics of making an interpretation run something like this: On the basis of his total background of experience and knowledge, an individual looking at a photograph is prepared to see many and sundry things. For example, if told that he is going to be shown a photo of a country scene, he will form some sort of mental image which includes trees and fields, perhaps hills, a farm and other sorts of things which he associates with the words country scene. If he knows something about the area covered he will have a more definite idea of what he will see in the photo. He will have a very strong expectation of seeing certain things. But for other things the expectancy may be so low that their appearance will be a complete surprise.

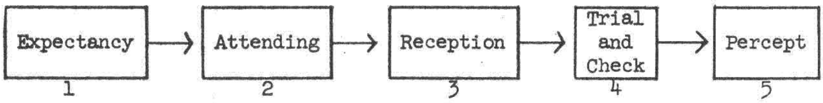
A military PI looking at photography of a section of the front, which he has seen before, may have a very high expectation of seeing trenches and gun emplacements because he has seen them there before. His expectancy would be somewhat less but still high if he had not previously seen photographic cover of this area. However, his expectancy of seeing coastal gun emplacements high in mountainous and the relatively inaccessible terrain would be extremely low.

The point is that the PI is *ready* or *set* to see certain things in a photograph. He carries in his head certain ideas about what he will see. These will depend on his knowledge of the specific situation and experience. The next step in this process is referred to as "attending," that is, looking at the photograph. He "attends to" or "looks at" certain things in the photograph more closely than others. If he expects to see certain kinds of things he will be predisposed to look at areas where he expects these things to appear. The third step is the reception of the visual image on the retina which is, generally speaking, a physical process though certain psychological factors are involved. This image is evaluated by the PI in the fourth step of this process by an implicit trial and check procedure. He guesses about the nature of the specific image. He may have to decide whether it is a tank, armored car or donkey cart. He attempts to substantiate each successive guess by finding additional information, such as track activity, separations in the shadow, etc. (Incidentally, in interviewing military PI's, it appears that the trial and check phase is very short. They quickly arrive at a final decision.) This last guess or final decision then is the last step in the process and is formally known as the formation of the "percept." It is the PI's interpretation of the specific image.

This study described in this paper concerns the first step of the process, that is, the "expectancy."

This factor of expectancy has been recognized for quite a long time under several different names such as set, unconscious inference, hypothesis, perceptual expectancy, to mention only a few. Its role has been stated as " . . . we perceive more readily and more clearly those events we expect to perceive." In fact, it has been found that if the expectation is sufficiently strong, distortions of perception occur to the extent of causing either seeing things that do not exist or a misidentification of other things. More generally, it

† The opinions expressed in this paper are those of the author and do not necessarily reflect official Department of the Army policy.



1. Expectancy: Things the PI expects to see in the photo.
2. Attending: The PI focuses his attention on a specific part of the photo.
3. Reception: The image is received on the retina.
4. Trial and Check: The PI evaluates the specific image.
5. Percept: The PI makes an interpretation of the image which he feels accounts most satisfactorily for the things he sees.

FIG. 1. A schema for the photo interpretation process.

may be said that *one tends to see what he expects to see.*

The effect of such a factor in military photo interpretation becomes especially important in view of the military PI often being given intelligence information that varies greatly in terms of both accuracy and reliability. If inaccurate intelligence information sets up expectancies in the PI which predispose him to make misinterpretations, or to produce wrong information, it will be necessary to find ways for compensating or correcting such erroneous information.

The problem may be stated as "How is a PI's interpretation affected by intelligence reports?"

Attempt was made to answer this question through a set of experiments in which school-trained military PI's were fed several kinds of intelligence information. Their interpretations of operational sets of photography were compared to those of matched groups of PI's who were not given such intelligence information.

These experiments were conducted as three separate exercises using sets of operational photography representing three kinds of situations: (a) a Korean Air Field, (b) Korean Countryside, and (c) a far Eastern Urban area. The subjects were supplied with the appropriate maps and overlays, response sheets, and specific information requests, together with appropriate background information detailing the friendly and enemy positions and activity. The experimental group was given additional information about enemy equipment and the information attributed to intelligence sources.

The subjects of this study were 21 officers and 38 enlisted men. They were administered these exercises during July and August 1960, immediately before graduation from the photo interpretation course at the Army Intelligence School, Fort Holabird, Maryland. Both the officers and the enlisted men were divided into two groups matched on the basis of final course grades and a composite of aptitude test scores. Each of the two matched groups served as the experimental group for one exercise, and as controls for one or both of the other two exercises. Table 1 describes the materials used, time for test administration, and outlines the design. The following is an example of one of the exercises:

"It is the month of July 1950, and you are a photo interpreter with the 25th Division PI team in Korea. In June and early July 1950, the Eighth U. S. Army is holding a large semicircle around PUSAN on the Southeast coast of Korea. Reports indicate heavy enemy troop movements to the southeast coast apparently aimed at PUSAN, the vital port through which most of our supplies funnel. In addition to their own equipment, the enemy is using captured U. S. equipment, including trucks, tanks, and aircraft.

"A request was received by your G2 Air for information on activities on the airfield near PO'HANG-DONG, recently captured by the enemy. Spot cover has been received at Eighth Army of many such areas along the coast, and the two aerial photos covering the airfield were forwarded to your team chief. He has determined that the scale of this photography is 1:4,000 and has placed the following requirements on you:

a. Locate and identify all tanks.
b. Locate and identify all heavy construction equipment.
c. Locate and identify all underground installations.
d. Locate and identify all components of the airfield defensive system.

TABLE 1

ADMINISTRATION OF PERFORMANCE MEASURES TO PI SAMPLES

Exercise	*No. of Photos*	*Photo Size*	*Scale*	*Administrative Time (Mins.)*	*Experimental (Added Intell. Info.)*	*Control Cond.*
Korean Air Field	2	9″×18″	1/4,000	75	Group A	Group B
Korean Countryside	2	9″×18″	1/5,700	15	Group B	Group A
Far Eastern Urban Area	11	9″×9″	1/15,000	20	Group B	Group A

Group A: Officer N=10, EM=19.
Group B: Officer=11, EM=19.

"You will have 75 minutes to complete this exercise.

"*Special Instructions:*

"Use only the names indicated on the accompanying object list in making your identifications. No further identifications will be required."

The object list details the specific objects for which the interpreter should search. For example, in this problem, the kinds of heavy construction equipment listed are bulldozers, cranes, power shovels, and road rollers, graders, and compacters.

The experimental group received exactly the same background information, requirements, and special instructions as were given to the control group. However, in addition, the experimental group received the following information:

"Agent reports received by your division G2 in the past two days indicate that the airfield is undergoing extensive repairs and that an infantry regiment supported by a tank battalion (T-34 tanks) has been moved in to provide defense to the airfield during this rebuilding period. A POW identified as a member of the 18th NKPA Engineer Battalion stated that large numbers (thirty to forty) of construction and earth-moving type equipment had been moved to the airfield by his unit."

Notice that the additional information directly suggests the presence of tanks and construction equipment. The presence of underground installations and airfield defenses is not specifically suggested, although it could be argued that their presence might be inferred from the known military situation.

The results of each of the three exercises were analyzed individually to determine whether introducing intelligence information resulted in differences in interpretations. The analysis of the data for the Korean Airfield problem indicated that the introduction of intelligence information had an effect on the PI's interpretations but was not clearly defined. In the requirement to find tanks for example, the experimental group, the group given the intelligence report, showed a stronger predisposition during the first 25 minutes and during the first 50 minutes to report tanks than was evidenced by the control group. A statistical test indicated that there was a real difference between these groups. However, by the end of the 75 minutes allowed for this exercise the number of persons reporting tanks tended to become more nearly the same for the two groups. Even though the experimental group continued to report more tanks, a statistical test indicated that the difference between the groups was too small to be accepted as significant.

Analysis of the reports of the presence of heavy construction equipment showed that the two groups did not differ significantly in reporting numbers of pieces of equipment.

In the Korean countryside problem, in which the subjects were asked to locate vehicles, the experimental group was given a report that a night reconnaissance patrol had seen vehicles being dispersed and camouflaged in the areas covered by the photography. The analysis of the responses made by the two groups indicated that the introduction of intelligence information did not create a sustained predisposition in the experimental group to perceive significantly more vehicles. During the middle 5 minutes of the exercise (this was a 15 minute exercise), a significant difference between the two groups did occur but it disappeared again by the end of the exercise.

In the urban area problem, the experimental group was told that a defecting municipal official had reported that unfriendly forces had placed military units in all hospitals in the area, and were using these hospitals as centers for political activity. He had also reported that this area was known for its many hospitals and that unfriendly forces had been using their control of these medical facilities to impose their will on the populace. In this exercise, both the experimental and control groups were required to search for all hospitals. In addition, both groups were required to search for all concentrations of 10 or more revetted buildings appearing in the photography. No suggestive information was given to either group as to the presence of revetted buildings however.

Analysis of the data—numbers of hospitals reported by the two groups—showed a significantly greater number of PI's in the experimental group reporting hospitals than in the control group for all time periods. No difference was found between the two groups in reporting revetted buildings.

To summarize these results, a *sustained* effect, due to the introduction of intelligence information, was found in only one case, that of hospitals. It is of interest to note that of all the objects used to test the effect of introducing intelligence information, hospitals are the most ambiguous and therefore, for the PI, are the most difficult to check. In some cases the intelligence information was effective in the initial time periods, and in others not at all effective.

In general, the data indicate that under some conditions the introduction of intelligence information has an effect on the PI performance, and that this effect is related to certain other variables of the photo interpretation situation. The variables hypothesized include such things as ambiguity of the object as noted in the case of hospitals, the effect of time as seen in the cases of tanks and vehicles, and object related factors. These variables will be investigated in future research as well as the general area of expectancy using experienced PI's.

Some methods which are now under study and which may compensate for the introduction of erroneous intelligence information, include such things as use of independent interpretations, peer and supervisor checks on interpretations, use of indications of confidence by the PI in his interpretation, and other somewhat similar devices. Other analyses which have been made of these data show that PI's are significantly more confident of their responses when they make a right response than when they make a wrong response. In other research which has been conducted it has been found that it is possible to substantially increase the proportion of right information, by pooling independent interpretations and by using items on which PI's agree.

Reference

Solley, C. M. and Murphy, G. *Development of the Perceptual World.* New York: Basic Books, Inc., 1960.

*Investigation of the Church Method for Orientation of a Single Aerial Photograph**

1/LT DONALD C. BURDITT,‡
544th Reconnaissance Technical Group,
Offutt Air Force Base, Nebraska

(*Abstract is on next page*)

Introduction

The Church Method is used to compute the tilt, swing, azimuth and exposure station coordinates for a single aerial photograph. The Alwac III-E computer program, which is currently being utilized for solution, was written by Charles W. Hanson, Broadview Research Corporation. The Church Method utilizes three or more ground-control points and determines the solution by an

* Publication in this Journal has been cleared by the Department of Defense.

‡ At the time of submitting this paper, the author was an Air Force officer. His new address is 7800 Seward St., Omaha, Neb.