Acta Psychologica 71 (1989) 23–51
North-Holland

UNSUPERVISED PRACTICE: THE PERFORMANCE OF THE CONTROL GROUP *

Mark A. FOSS, Monica FABIANI, Amir M. MANÉ
and Emanuel DONCHIN

University of Illinois at Urbana-Champaign, USA

A control group of 40 subjects practiced the Space Fortress game for 10, one-hour, sessions. They were given standard game instructions, but were not aided in their training in any other fashion. Subjects in this group showed a general improvement, throughout training, in the total game score as well as in many other aspects of game performance. However, individual differences were found in the subjects' initial capability, in their rate of learning and in the strategies they adopted to achieve their final performance. In order to summarize the many aspects of this complex database, two multivariate techniques were used: Three-Mode Principal Component Analysis and Cluster Analysis. These techniques proved useful in that they provided a coherent and relatively simple description of the subjects' behavior. The model derived from these multivariate procedures was applied to an independent group of subjects. This cross-validation accounted for some of the differences observed between the two groups.

Introduction

Several groups of investigators collaborated in the Learning Strategies project to examine the extent to which it is possible to improve on

* The research described was supported by DARPA Research Contract #MDA903-84-K-0065, Dr. Craig Fields contract monitor, Lt. Colonel Jack A. Thorpe Scientific Program Manager. The research described was also supported by ARI Research Contract #MDA903-86-C-0346, Dr. Michael Kaplan director.

We acknowledge the contribution of Russel Sheptak in the collection of the data, of Brian Foote, Earle Heffley and William Sherman in the development of software, and of Mike Anderson and Walter Meyers for technical assistance. We are indebted to Jean Buckley and Gabriele Gratton for helpful comments on earlier versions of this manuscript.

Mark A. Foss is currently at the Dept. of Psychology, Temple University, Philadelphia, PA 19122, USA.

Requests for reprints should be sent to: E. Donchin, Psychology Dept., University of Illinois, 603 East Daniel, Champaign, IL 61820, USA.

unsupervised practice in the acquisition of complex skills (see Donchin 1989, this volume). All the investigators assessed training in the context of the Space Fortress task and each manipulated some aspects of the task in order to examine their effects on performance. It was necessary to provide a common baseline against which the experimental results could be compared. We assembled, therefore, a control group of 40 subjects, who practiced the Space Fortress game for a period of 10, one-hour, sessions. These subjects were given instructions concerning the rules of the game at the beginning of the training period. During the training period, they practiced the Space Fortress game without any additional instruction or feedback. Thus, their training was 'unsupervised', in that the choice of game strategies was left entirely to the subjects' own devices, and the sole role of the experimenter was to monitor the subjects' progress.

As was true for all the projects described in this volume, the prime comparison between subjects was made in terms of a composite, or total, score which combined various aspects of the subjects' performance. All the subjects were informed of the contributions of various game events to the total score and were instructed to attempt to maximize it. Thus, the learning curves showing the value of this score as a function of the time in training served as the official comparison measure across projects. However, we also maintained a very detailed record (which included over 200 dependent measures) of the subjects' performance throughout the training sessions (Mané and Donchin 1989, this volume). An examination of these data revealed that similar scores could be reached by various routes as subjects varied their strategies. However, the massive size of the data base made it very difficult to describe and interpret the nature of these individual differences. Therefore, we reduced the dimensionality of the data base by means of Three-Mode Principal Component Analysis (Three-Mode PCA, Tucker 1966). This report describes the data obtained from the control group both in terms of the standard learning curves used in other papers in this volume and in terms of the descriptive space generated by the Three-Mode analysis. In addition, to illustrate the utility of that space we also examined, within its framework, the data obtained from the control group run by Gopher et al. (1989, this volume).

Method

Subjects

Forty male students at the University of Illinois were paid $3.50 per hour to participate in the study. They were all right-handed, 18–24 years of age, and had normal or corrected to normal vision. They were selected from a larger pool of subjects on the basis of a score obtained in an *aiming screening task* (Mané 1985; Mané and Donchin 1989, this volume; Mané et al. 1989, this volume). A minimum aiming screening score of 780 points was required for the subject to participate in the study. A description of the aiming screening task and of statistics concerning the screening scores are presented in the Results section.

Procedure

Subjects were trained to play the Space Fortress game (for a description, see Mané and Donchin 1989, this volume) over a period of 10 sessions, each lasting approximately one hour. Sessions were run on consecutive days, with no more than a two-day interruption between sessions. During the first session, subjects first received videotaped instructions illustrating all the rules of the game with the exception of those concerning the management of resources. Then they played a practice game block (5 minutes) in which the resource option was turned off. After this practice game, subjects were given instructions about the use of resources, and played four additional game blocks in which the resource option was reinstated (standard Space Fortress game). During the second session subjects played the standard Space Fortress game for 7 blocks (each lasting 5 minutes), while from the third through the tenth session they played 8 games daily. Thus, subjects practiced the Space Fortress game for a total of 76, 5-minute, blocks. However, as the first game block was different from all the others, only 75 blocks are included in the analyses reported in this paper. [1]

Results

Learning curves

Average learning curves for the group of 40 subjects are shown in figs. 1 and 2. The subjects' performance in the Space Fortress game appears to improve steadily throughout the training period. This can be seen with respect to the total score (fig. 1) as well as with respect to a number of other dependent measures. [2] For example, the number of fortresses destroyed by the subjects, which is a key contributor to the total

[1] Note that the same training schedule was used for all the control groups run by the individual contractors.

[2] These measures are described in detail in Mané and Donchin (1989, this volume).

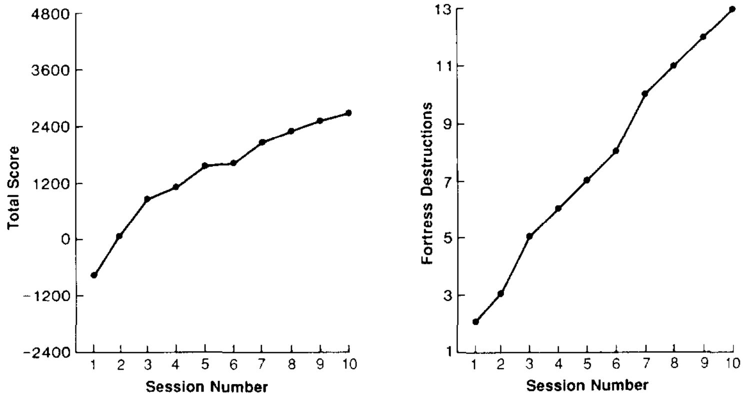

Fig. 1. Learning curves for the total game score and the number of fortress destructions. The training session number is plotted on the abscissa. The dependent variables are plotted on the ordinates. The average values for each session are plotted.

score, also increases (fig. 1). The number of shells fired by the fortress at the ship (fig. 2a), the number of times the ship crosses the line-of-fire of the fortress and thus risks being hit (fig. 2b), the number of times the ship wraps around the screen (fig. 2c), the mean reaction time of foe mines (fig. 2d) and its standard deviation (fig. 2e), and the number of IFF intervals outside the permitted range (fig. 2f) all decrease over the course of training. [3] It is interesting to note that performance does not appear to have reached its asymptote on any of these dependent measures. However, as will be shown in the next section, individual subjects deviate from these average trends in a number of ways.

Individual differences

Starting capability: The aiming screening task

In previous research (Mané et al. 1984), the aiming task was developed as a screening and placement tool, because the subject's success in the aiming screening task correlates positively with the success of the subject in the whole game. In the aiming screening task the spaceship can only rotate. Every second, a stationary mine appears somewhere on the screen. The subject has to aim and shoot at the mine. For every hit the subject gets 20 points and his final score reflects the number of mines destroyed in a 2-min block. Subjects performed the aiming screening task three times, and the best score they obtained was used to estimate their initial capability (see also Mané and Donchin 1989, this volume).

To establish a pool of subjects, 101 students were paid $1 to perform three blocks of the aiming screening task. The parameters of the distribution of the screening scores is

[3] Note that for all the dependent measures presented in fig. 2, a decrease represents a bettering of performance.

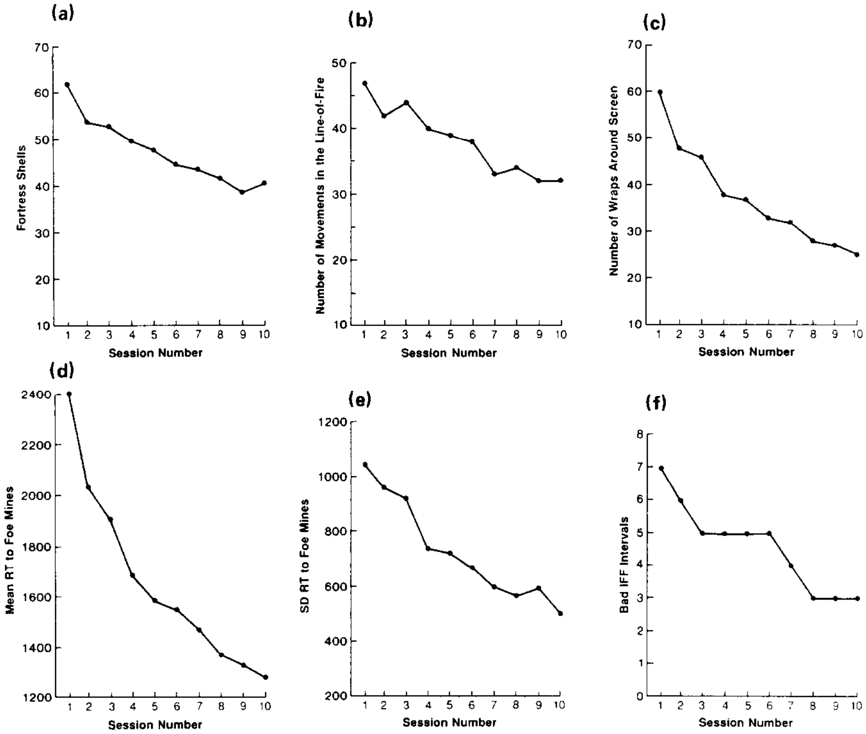

Fig. 2. Learning curves for a series of variables related, respectively, to the pursuit behavior of the fortress, to the movement of the ship, and to the handling of mines. The training session number is plotted on the abscissa. The dependent variables are plotted on the ordinates. The average values for each session are plotted. Note that for all these variables a negative trend of the learning curves represents an improvement in performance.

presented in table 1. Subjects who scored below 780 points were not invited to participate in the experiment. [4] In addition, based on these results, five categories of predicted success were defined. The range of scores for the five levels are presented in table 2.

The average screening score for the 40 subjects in this study was 1027.5 with a standard deviation of 129.0 (maximum score of 1280). Fig. 3a depicts the relationship between the aiming screening task scores of these subjects and their average total score in the Space Fortress game during the 10th training session ($r = 0.4419$, $p < 0.01$). Fig. 3b depicts the relationship between their aiming screening task score and the average number of fortress hits in session 10 ($r = 0.6687$, $p < 0.01$). It appears that the aiming

[4] This cut-off score was sufficiently low to guarantee a wide range of initial capabilities, and sufficiently high to allow the exclusion of those subjects whose rate of learning would be too slow to be studied in a reasonable amount of time.

Table 1
The distribution of the aiming screening score data.

Number of values	= 101
Minimum value	= 380.00
Maximum value	= 1340.00
Mean	= 990.10
Standard deviation	= 171.95
Skew	= −0.91
Kurtosis	= 1.07

Table 2
The five levels of expertise determined on the basis of the aiming screening task.

	Range	N	ULN	Available N
(a)	780–920	(19)	(5)	(24)
(b)	920–1000	(18)	(5)	(22)
(c)	1000–1060	(16)	(4)	(20)
(d)	1060–1160	(18)	(6)	(24)
(e)	1160+	(21)		(21)

Note:
N = the number of subjects in each level (the score of 780 is the cutoff point).
ULN = Upper limit N. It includes subjects that could be assigned to that group or to the group immediately above. This allows more flexibility in the assignment.
Available N = Available N is the N of the group plus the upper limit N.

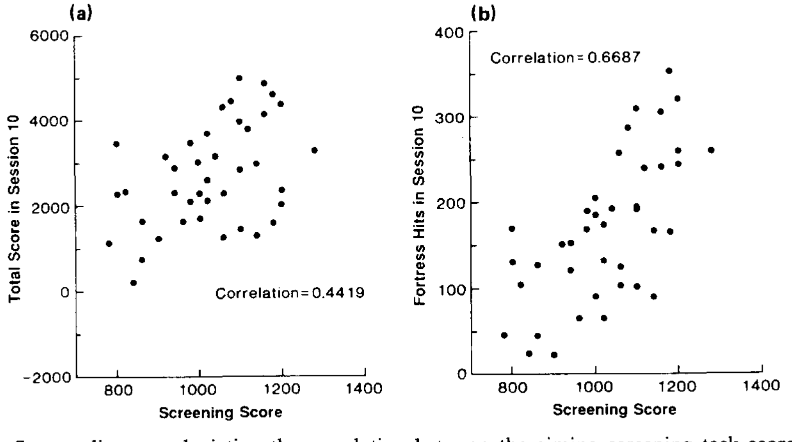

Fig. 3. Scatter diagrams depicting the correlation between the aiming screening task score and, respectively, the total score and the number of fortress hits in session 10. The screening score is plotted on the abscissa and the dependent variables on the ordinate.

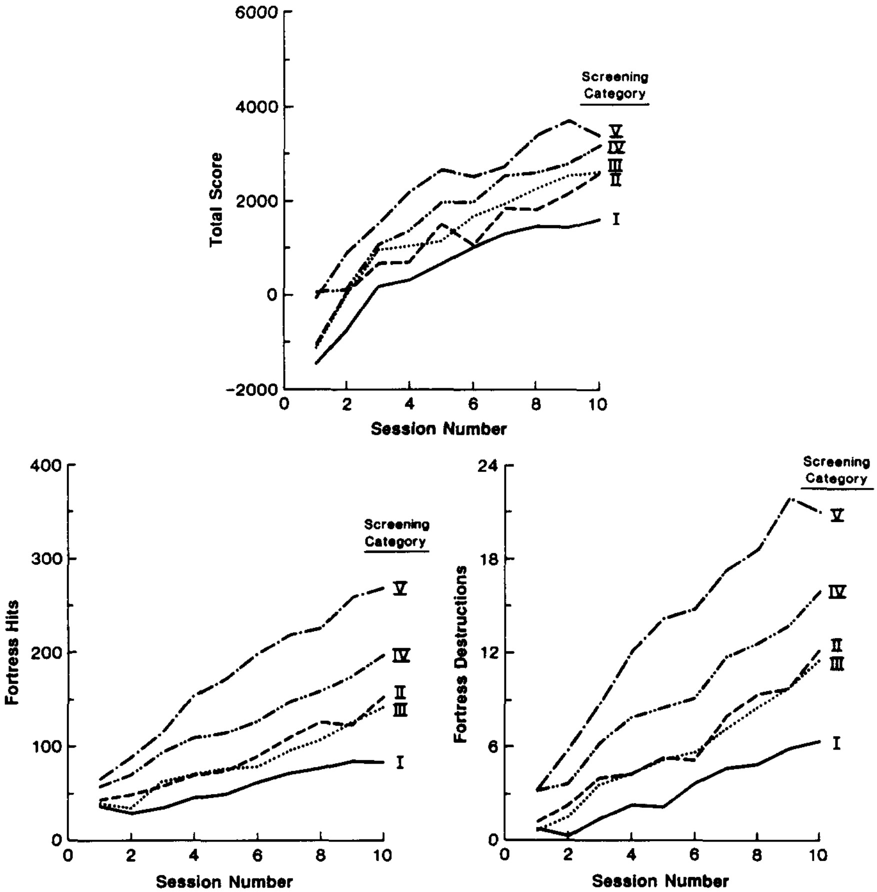

Fig. 4. Learning curves for the total game score, the number of fortress hits, and the number of fortress destructions, for subjects belonging to different screening categories. The training session number is plotted on the abscissa. The dependent variables are plotted on the ordinates. The average values for each session are plotted. Note that level V includes subjects with the highest aiming screening scores, and level I subjects with the lowest screening scores.

screening task score was a good predictor of the final overall performance of the subjects in this study, and an even better predictor of some of the aspects of the game related to aiming and shooting efficiency.

Fig. 4 shows the learning curves of subjects that were assigned to different screening categories for three of the dependent variables: total score, number of fortress hits and number of fortress destructions. It appears that the individual differences determined on the basis of the aiming screening task are manifest at the beginning of training, and are clearly maintained, if not increased, over time.

Table 3
Ship maneuvering strategies.

sub	cluster	aim	sumx	sumy	stx	sty	wrap	movin	score	self report
(1a) Slow circling (N = 7)										
12	3	1100	306	304	860	321	2	2	5267	circling c
14	3	1160	409	419	1103	562	6	5	4910	circling
17	3	1200	446	510	1289	573	21	13	4186	drift slowly
18	3	1200	431	455	1430	601	11	9	5508	circling
31	2	1040	395	427	1075	550	10	11	3572	circling
21	3	1280	359	630	1270	239	42	16	4910	diag ln slow
34	2	1160	552	609	1317	796	20	12	4534	circling
mean		1163	428	479	1192	520	16	10	4699	
(*sd*)		(77)	(74)	(114)	(191)	(186)	(13)	(5)	(661)	
(1b) Slow circling with more joy-stick manipulation (especially on Y) (N = 4)										
35	2	980	661	700	1858	1671	27	26	3172	circling
39	6	940	663	901	2081	1530	47	41	2868	circling
6	2	1100	711	770	2041	1834	21	20	3355	circling
29	2	800	742	746	2114	1822	18	18	3494	circling c
mean		955	694	779	2024	1714	28	26	3222	
(*sd*)		(124)	(39)	(86)	(114)	(143)	(13)	(10)	(270)	
MEAN ($N = 11$)		1087	516	588	1494	954	20	16	4161	
(*SD*)		(138)	(155)	(182)	(449)	(624)	(14)	(11)	(916)	
(2a) Rapid circling (N = 10)										
1	1	1140	834	856	3583	3840	20	20	3376	circling
11	6	1000	872	893	2943	3917	28	34	2615	circling
16	6	1140	819	879	2437	2887	29	34	1960	circling c
19	1	1120	859	878	3296	3567	20	24	3996	circling cc
23	1	1180	807	841	3131	3091	17	29	5139	circling

33	1	1200	823	987	3345	3669	33	31	4498	circling
37	4	1000	818	905	2130	2072	34	24	2357	circling
38	4	1060	856	903	2799	3155	24	34	2124	circling cc
7	6	1100	821	871	2431	2978	40	42	2200	circling cc
26	5	1020	883	825	2093	2257	54	48	2582	circling
mean		1096	839	884	2819	3143	30	32	3085	
(*sd*)		(73)	(26)	(45)	(527)	(630)	(11)	(8)	(1111)	
(2b) Rapid circling with very high level of stick movement (N = 5)										
5	1	1080	836	895	3889	4643	13	44	4414	circling c
20	1	800	895	886	4412	4787	10	33	2944	circling c
10	1	1060	891	926	4211	5070	15	36	4828	circling
28	1	1000	855	931	3805	5088	11	39	3499	circling cc
22	6	860	844	904	3301	5334	23	72	1948	circling
mean		960	864	908	3924	4984	14	45	3527	
(*sd*)		(124)	(27)	(20)	(425)	(272)	(5)	(16)	(1153)	
MEAN ($N = 15$)		1051	848	892	3187	3757	25	36	3232	
(*SD*)		(111)	(28)	(39)	(722)	(1041)	(12)	(12)	(1104)	
(3) Straight-line flight (N = 8)										
2	7	1100	127	1413	1256	673	73	66	4028	line
4	7	1020	352	1086	1226	650	61	53	2620	line
9	7	940	120	1393	1227	317	71	68	2289	line
13	7	980	81	1357	1645	247	68	65	3645	line
15	7	1020	110	1394	1060	256	72	67	3369	line
24	7	920	295	1267	1501	487	73	61	3154	hover
25	7	960	215	1235	1120	335	68	58	2564	line, circle
30	7	840	92	1358	899	158	70	65	2575	line
MEAN ($N = 8$)		973	174	1313	1242	390	70	63	3031	
(*SD*)		(78)	(102)	(111)	(238)	(192)	(4)	(5)	(615)	

[*Continued on following page*]

Table 3 (continued)

sub	cluster	aim	sumx	sumy	stx	sty	wrap	movin	score	self report
(4) Unidentified (N = 6)										
8	6	1180	782	914	2200	1328	63	43	2342	circling
27	5	780	756	806	1549	943	72	39	1666	Unspecified
32	6	1060	926	826	1991	1602	74	52	2320	line (diag?)
40	4	860	836	949	2773	4360	50	48	2003	circling
36	4	900	534	1111	1420	4946	60	64	1845	line, circle
3	4	820	750	978	2523	3545	34	47	2510	circling c
MEAN ($N = 6$)		933	764	931	2076	2787	59	49	2114	
(*SD*)		(155)	(130)	(111)	(532)	(1711)	(15)	(9)	(328)	

Note:
sub = subject number (1–40),
cluster = cluster to which the subject belongs in the Cluster Analysis,
aim = score in the aiming screening task,
sumx = ship movement on the screen in the X-axis,
sumy = ship movement on the screen in the Y-axis,
stx = joy-stick movement in the X-axis,
sty = joy-stick movement in the Y-axis,
wrap = number of times the ship flies off the screen,
movin = number of times the ship crosses the line-of-fire of the fortress,
score = highest score obtained by the subject in session 10,
self report = self report of the subject's flight strategy, collected at the end of training (c = clockwise, cc = counter-clockwise, ln = line, diag = diagonal).

Ship maneuvering strategies

The instructions for the Space Fortress game included only a brief reference to the optimal pattern of maneuvering the spaceship. A suggestion was made to avoid flying in a straight line. This instruction set allowed the subjects to develop individual flying patterns. Monitoring the results of the control group, there were three general categories of flight patterns: slow circling, rapid circling and straight-line flight. The classification was based on the measurement of the dependent variables which are relevant to the control of the ship, obtained throughout the 10 hours of training. The individual scores fell rather neatly into these three patterns and only about 15% of the subjects did not fit well in any of the categories.

(a) Slow circling. Subjects using this flight pattern had low ship movement on the X and Y dimensions with about the same amount of movement on both dimensions. There were very few occasions where the subject wrapped around the screen and few movements into the line-of-fire of the fortress. There was relatively little manipulation of the joy-stick. This group can be segmented into two subgroups on the basis of the joy-stick manipulation (especially on the Y dimension). The subgroup with low stick manipulation had the highest game score. However, this group also had a better screening score. Data for this group are presented in table 3.

(b) Rapid circling. These subjects flew around the screen with relatively high speed. They had a very high level of manipulation of the joy-stick in both the X and the Y dimensions. The result was a similar high amount of movement on both dimensions on the screen. They were intermediate with respect to the number of wraps around the screen and the movements into the line-of-fire of the fortress. Data for this group are presented in table 3. This group can also be subdivided into two subgroups, again on the basis of stick movement.

(c) Straight-line flight. The subjects in this group forfeited the control of ship acceleration. They chose to give the ship an initial acceleration and from that point on mostly controlled its rotation. They aimed at the fortress and readjusted their aiming as they flew across the screen. This flight pattern involved little manipulation of the joy-stick on the Y dimension and more on the X dimension. The amount of ship movement on the X dimension of the screen was low and this dimension discriminates clearly between the subjects who flew in a straight line and those who did not. The ratio of movement on the Y dimension over movement on the X dimension was as high as 10 : 1. This flying pattern also resulted in a high number of wraps around the screen. Data for this group are presented in table 3.

(d) Unidentified flight pattern. Six subjects were excluded from the above classification. As already mentioned, the values used for classification were based on performance over the 10 hours of training. The subjects included in this group may be those whose strategies changed over the course of training. For them, a classification that uses only the average of the last session might yield better results. Data for these subjects are presented in table 3.

Figs. 5–7 show the learning curves for subjects using each of the flight strategies (unidentified subjects were excluded). Subjects in the slow-circle group appear to obtain better game scores throughout training than the other two groups (fig. 5). The

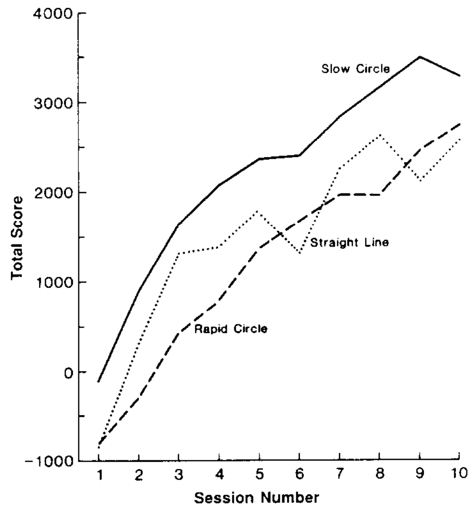

Fig. 5. Learning curves for the total game score for subjects using different ship maneuvering strategies. The training session number is plotted on the abscissa. The total score is plotted on the ordinate. The average value for each session is plotted.

measure of movement on the X-axis clearly discriminates between the three groups, with the rapid-circle group at the top and the straight-line group at the bottom (fig. 6a). The number of ship movements into the line-of-fire of the fortress also discriminates among the groups: the straight-line group has the highest number (as they cross the line-of-fire of the fortress every time they wrap around the screen), and the slow-circle group the lowest (fig. 6b). The number of wraps around the screen discriminates between the subjects using a straight-line strategy and the other subjects (fig. 6c). Fig. 7 shows that, even though the strategies of the groups are clearly distinguishable on the basis of movement-related variables, they are not as clearly distinct when variables related to mine handling are plotted. The slow-circle group appears to be slightly but consistently faster in destroying foe mines (fig. 7a). However, this is not due to the fact that their reaction times to the identification of foe mines (fig. 7b) are faster than those of the groups.

The strategy classification presented here for the control group provides a useful way of comparing the individual differences observed by other investigators. For example, it seems that one of the effects of giving the subject more specific guidance in the training period is a reduction of individual differences in flight strategies among subjects trained with the same method (see Fabiani et al. 1989, this volume, for an example). However, there is at least one major drawback in the subject classification presented here, namely, that it is based solely on the dependent variables related to the maneuvering of the ship. Subjects could be characterized also on the basis of their total score, or on any of the many other dependent measures recorded, or on the basis of how rapidly the subjects improve on each of these measures over time. It is because of these reasons that, in the second phase of this study, we attempted to take all of these

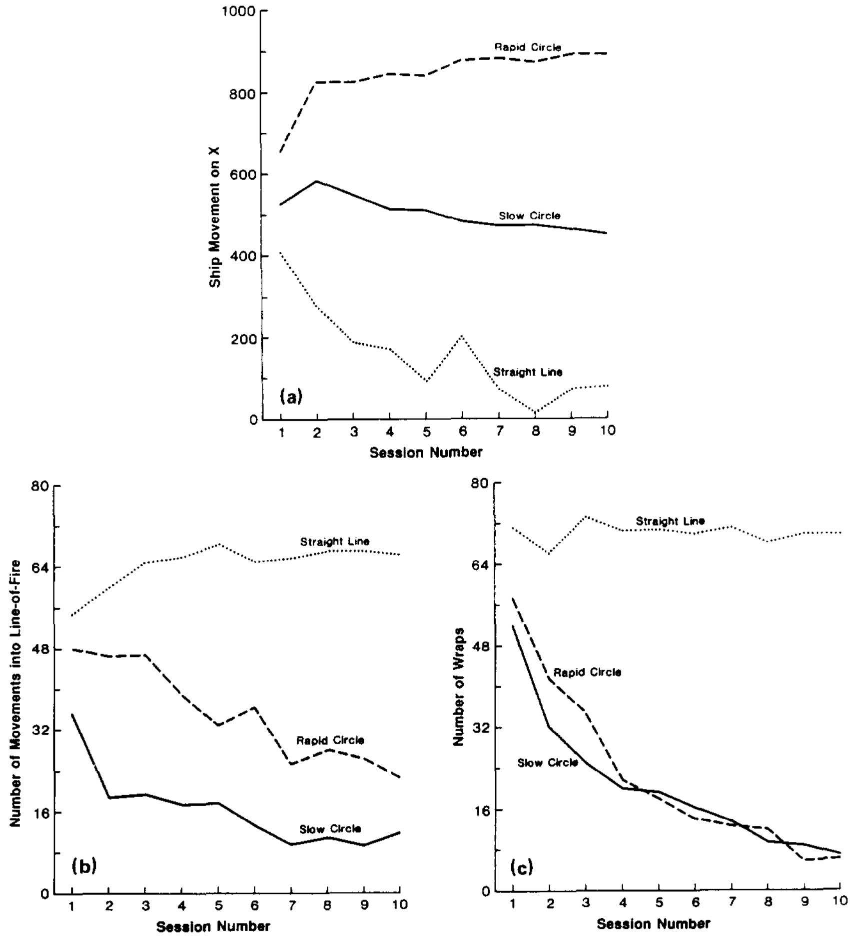

Fig. 6. Learning curves for a series of variables related to the movement of the ship for subjects using different ship maneuvering strategies. The training session number is plotted on the abscissa. The dependent variables are plotted on the ordinates. The average values for each session are plotted.

factors into consideration by analyzing the data with two multivariate techniques – Three-Mode PCA and Cluster Analysis – that are described in the following section.

Three-Mode Principal Component Analysis

Rationale

Because individuals differ, not only in performance but also in their learning rate and in their style of play or strategy, a description of learning in the Space Fortress

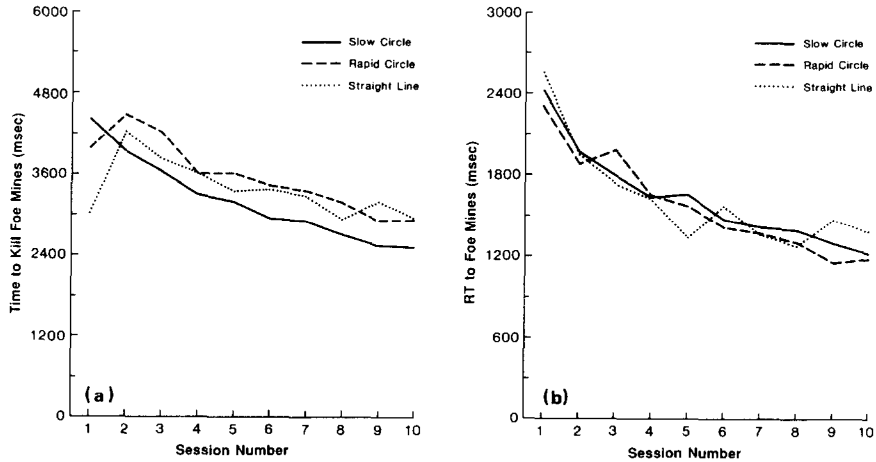

Fig. 7. Learning curves for two variables related to mine handling for subjects using different ship maneuvering strategies. The training session number is plotted on the abscissa. The dependent variables are plotted on the ordinates. The average values for each session are plotted.

game can be represented in three statistical 'spaces' or 'modes' – subjects, time, and measures of performance used.

One method for creating a simplified description of how individuals differ in their strategies and learning rate over time is Three-Mode PCA (Tucker 1966). The Three-Mode PCA model is, in essence, the principal components of each mode (subjects, measures, and time-blocks) along with a matrix of interactions between modes.

Specifically, the Three-Mode PCA model seeks a parsimonious description of the data by finding a sufficient set of principal components for each mode of the data such that the product of the component loadings, weighted by an interaction term, yields a least squares estimate of the data (see Kroonenberg 1983, 1984). The interaction terms are measures of the relationship between the components from different modes (called by Tucker the 'core matrix'). The model is expressed by the following equation:

$$z_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} g_{ip} h_{jq} e_{kr} c_{pqr} + r_{ijk}, \tag{1}$$

where:

Z ≡ the three-mode Data Matrix of Subjects [$i = 1, 2, \ldots, l$], Measures [$j = 1, 2, \ldots, m$], and Blocks [$k = 1, 2, \ldots, n$];

G ≡ the Subject Loading Matrix [l by s];

H ≡ the Measure Loading Matrix [m by t];

E ≡ the Block Loading Matrix [n by u];

C ≡ the Core Matrix of Component Interactions [s by t by u];

R ≡ the Residual Matrix [l by m by n].

A large value for c_{pqr} indicates that the particular combination of components, one from each mode, is predictive of individuals' scores. That is, the subject component p during the block component q for the measure component r is predictive of z_{ijk} when i, j, and k load heavily on their respective components (i.e., g_{ip}, h_{jq}, and e_{kr} are all large). Since the component loadings are multiplied together, the combination contributes in predicting the score only if all of them are non-zero. For example, if a particular combination predicts most subjects' scores on a measure during a particular block but does not do so for subject i, then the subject loading – g_{ip} – would be low. That is, this particular combination does not explain this subject's performance.

Measures

Twenty-one measures of game performance were used for all the following analyses (see table 4). These measures were designed to assess five facets of game play: (a) *movement of the ship* vertically, horizontally, into hyperspace (off the screen and re-entering on the opposite side), and into the fortress' line-of-fire; (b) measures of *tactical effectiveness* such as frequency of ship damage, damage to the fortress, time taken to destroy the fortress, and percentage of shots that hit their target; (c) measures related to *the time required for mine (friend and foe) recognition and destruction/energizing*; (d) measures related to the *initiation and completion of the IFF double press (foe mines)*; [5] and (e) *resource management* measures (indicating the effective use of the bonus and missile options).

Some of these measures could not be computed for some of the games because their computation would involve a division by zero. These occurrences were handled in the manner outlined in the following examples. If a subject failed to destroy the fortress during a game (median occurrence less than four blocks per subject) then the variable which indicates the average time taken to destroy the fortress (AVKTFT) was set to its largest value plus one. If the subject failed to initiate a double press of the IFF button during a block (five subjects had 1 to 4 such games) the mean reaction time and its standard deviation from the appearance of the mine to the first IFF press (MEAN RT and STDEVRT), and the mean interval between the first and second IFF press and its standard deviation (MINTERV, and STDEVINT) were set to the average of the three preceding and two following scores when they were available. Finally, if the subject failed to destroy any foe mines during a game (median occurrence was less than once per subject) than the average time taken to destroy foe mines (AVKTFO) was set to the subject's average value on this variable across blocks.

Analyses and results for the Illinois sample

To summarize, the data consisted of 21 measures on 75 blocks (games) from 40 subjects. The observations on each measure were standardized across subjects and games and then analyzed via a Three-Mode PCA (Tucker 1966) using Kroonenberg's

[5] Whenever a foe mine appears on the screen, the subject is required to switch weapon system. To do that, he has to press twice a button designated as IFF button (Identification Friend or Foe), with an interval between 250–400 ms between button presses.

Table 4
The dependent measures of game performance.

Ship movement	
WSUMVX	Velocity of the ship in the X-axis.
WSUMVY	Velocity of the ship in the Y-axis.
STSUMX	Number of loops in which motion in the X-axis was called for.
STSUMY	Number of loops in which motion in the Y-axis was called for.
NWRAP	Number of times the ship moved through hyperspace.
NMOVIN	Number of times that the ship moved into the fortress' line-of-fire.
Tactics	
$NSHDMG_1$	The number of times the ship was damaged by a mine.
$NSHDMG_2$	The number of times the ship was damaged by the fortress.
NFOHITS	Number of times the fortress was hit.
AVKTFT	Average time it took to destroy the fortress.
SHOTEFF	The percentage of shots that hit the target.
Mine handling	
MEAN RT	Mean reaction time to the identification of a foe mine (i.e., from the appearance of the mine to the first IFF press).
STDDEVRT	The standard deviation of the reaction time.
AVKTFO	Average time it took to destroy a foe mine.
AVKTFR	Average time it took to energize a friendly mine.
IFF timing	
MINTERV	The average IFF interval.
STDEVINT	The standard deviation of the IFF interval.
NBDINT	Number of bad IFF intervals.
Resource use	
PRCNTBON	The percent of the time the bonus was used.
RESMANAG	How advantageous was the use of the bonus option.
TOTSHOTS	The total number of shots.

alternating least squares program (Kroonenberg and De Leeuw 1980; Kroonenberg and Brounwer 1985).

A solution was selected that contained six subject components, four block components, and six measure components. The subject components explained 65% of the subject mode's variance, the measure components explained 77%of the measure mode's variance, and the block components explained 77% of the block mode's variance. The total model explained 60% of the total variance. [6] The number of components retained in this solution was determined by our ability to interpret the dimensions for each mode and by the total amount of variance accounted for by the model.

[6] This value is always less than the smallest individual mode value.

In an effort to make these results easier to interpret the principal components from each mode were rotated. What was sought was a set of dimensions that have simple structure. [7] It was found that for both the block and the measure components a VARIMAX rotation (Kaiser 1958) yielded a relatively simple structure that was easy to interpret.

A careful consideration of the role of the subject components in the Three-Mode PCA model revealed that these components are unlikely to have a simple structure. A subject's loadings on the various components represent the degree to which particular sets of measures during certain blocks of game are indicative of that subject's behavior. For example, one subject component could be related to the subject's tactical effectiveness in the early game blocks while another could correspond to the amount of straight-line flight in the later game blocks, etc.. A subject's behavior would then be described by a composite of all of these various measure–block interactions. Therefore, seeking a simple structure (i.e., where subjects load on one and only one dimension) is unreasonable. Any common patterns that subjects (or subsets of subjects) have in their measures–block interactions will be found in the core matrix. To the extent that there are similar patterns in behavior across subjects this should produce a simple structure in the subject facet of the core matrix. Therefore, a rotation was obtained for the subject facet of the core matrix and then the inverse rotation was applied to the subject components. Because there was no reason to expect an orthogonal structure, a rotational procedure was used that permits an oblique rotation if needed (DAPPFR, Tucker and Finkbeiner 1981).

The rotated block components simply divided the 75 blocks into four consecutive stages. The first component consisted of the first 5 blocks, the second consisted of the next 14, the third of the next 28, and the last consisted of the final 28 blocks.

The loadings for the rotated measure components are given in table 5. The first measure component indicates *tactical ineffectiveness*, such that a high score on this component represents frequent ship damage, poor shooting efficiency, and under-use of the bonus option. A high score on the second measure component indicates *straight-line flight through the Y axis*. A high score on the third measure component represents *conservative tactics*, such that few shots are taken resulting in infrequent hitting of the fortress and longer time to destroy it, but good use of the bonus option (i.e., the bonus points are appopriately chosen given the low utilization of missiles). The fourth measure component represents *foe mine identification*, such that a high score on this component indicates a long time before initiating the identification of foe mines and numerous bad double presses (i.e., the IFF interval was too long or too short) resulting in a delay in destroying the foe mines. The fifth measure component corresponds to *interval timing*, such that a high score on this component indicates that the IFF intervals tended to be long and variable. The sixth measure component represents the level of *stick control*, that is, a high score on this component indicates high degree of joy-stick manipulation on both the X and Y dimensions and high degree of movement of the ship on the screen on the X-axis.

[7] A simple structure is obtained when highly interrelated elements in a mode (e.g., the measures having to do with controlling the ship's movement) load heavily on only one dimension.

Table 5
Measure components.

Measures	Measure components					
	MC1	MC2	MC3	MC4	MC5	MC6
$NSHDMG_1$	*0.44*	−0.01	0.04	−0.07	0.07	0.19
$NSHDMG_2$	*0.44*	−0.13	−0.02	−0.02	−0.09	−0.00
SHOTEFF	*−0.48*	−0.12	−0.05	0.02	0.03	0.02
PRCNTBON	*−0.27*	−0.11	0.14	−0.23	0.00	0.09
WSUMVY	−0.13	*0.62*	0.04	−0.05	0.02	0.10
NMOVIN	0.08	*0.57*	−0.00	−0.05	−0.02	0.07
NWRAP	0.05	*0.38*	0.03	0.16	−0.01	−0.21
RESMANAG	−0.25	−0.09	*0.51*	−0.15	−0.03	0.08
AVKTFT	0.14	0.07	*0.28*	−0.12	0.08	−0.08
AVKTFR	0.22	−0.02	*0.23*	0.14	−0.01	0.00
NFOHITS	−0.14	−0.09	*−0.47*	−0.07	0.05	0.01
TOTSHOTS	0.08	0.02	*−0.57*	−0.13	−0.01	0.01
NBDINT	−0.25	0.06	−0.10	*0.64*	−0.05	0.06
AVKTFO	−0.02	0.02	0.09	*0.52*	−0.01	0.06
MEAN RT	0.22	−0.07	−0.00	*0.25*	0.15	−0.08
STDDEVRT	0.15	−0.06	0.07	*0.22*	0.09	−0.05
MINTERV	−0.04	−0.00	0.01	−0.06	*0.81*	0.01
STDEVINT	−0.01	0.00	−0.02	0.08	*0.54*	0.01
STSUMX	0.01	0.10	−0.12	−0.02	0.03	*0.59*
STSUMY	−0.01	0.06	0.05	−0.06	−0.01	*0.57*
WSUMVX	0.09	−0.22	0.04	0.19	−0.03	*0.44*

Note:
MC1 – Tactical ineffectiveness.
MC2 – Straight line flight (Y-axis).
MC3 – Conservative tactics.
MC4 – Mine identification.
MC5 – Interval timing.
MC6 – Stick control.

The subject mode is only interpretable in conjunction with the core matrix of component interactions. Furthermore, caution needs to be taken in that a subject component corresponds only to a structurally coherent aspect of subjects' behavior and *not* necessarily to some 'idealized' subject (cf. Tucker and Messick 1963; see also Cliff 1968). Each subject is a composite of many, and sometimes all, of these subject components. An alternative approach to examining the subject components is offered in the Cluster Analysis section presented below. In table 6 is the $6 \times 6 \times 4$ (subject $\times$ measure $\times$ block) core matrix.

Of the six subject components, one appears to be a general component, while the other five are specific to different measure components. The general subject component is *component 2*, which is composed of measure components 1, 3, 4, and 5. In each case the distinction starts out positively on the measure component and then shift to a negative distinction in the later blocks. Because all the subject loadings on this

Table 6
Core matrix.

Measures	Trial	Subjects					
		SC1	SC2	SC3	SC4	SC5	SC6
MC1	TC1	4.5	*52.4*	−0.1	1.6	2.6	*4.2*
	TC2	0.0	*36.7*	4.4	2.8	−1.7	*21.9*
	TC3	−1.2	*3.8*	1.0	−0.1	−1.0	*36.2*
	TC4	2.1	*−32.8*	−0.4	5.4	−1.7	*28.3*
MC2	TC1	*6.8*	13.5	0.4	−0.7	−2.4	3.1
	TC2	*31.9*	14.1	3.3	−0.6	−0.2	0.8
	TC3	*49.9*	5.8	0.5	3.5	0.7	−0.9
	TC4	*57.5*	−5.9	−2.8	−2.6	1.1	0.2
MC3	TC1	0.2	*3.2*	4.0	3.7	*8.7*	−5.5
	TC2	−2.4	*5.0*	*8.6*	*4.2*	*24.6*	−14.4
	TC3	1.3	*−2.0*	0.7	−3.0	*48.7*	−3.4
	TC4	3.3	*−26.4*	−2.7	3.4	*50.3*	4.7
MC4	TC1	−1.5	30.1	−2.8	*4.0*	−1.9	−2.9
	TC2	−1.9	*19.3*	−1.2	*16.5*	2.6	1.6
	TC3	−0.8	*−1.1*	1.0	*44.3*	4.2	0.4
	TC4	0.9	*−33.8*	1.7	*31.4*	5.6	3.0
MC5	TC1	−3.2	*17.7*	−9.0	−3.6	10.1	4.1
	TC2	1.1	*−0.1*	−3.1	53.0	−6.0	−24.2
	TC3	−0.5	*−2.6*	−3.2	−3.2	2.9	4.1
	TC4	0.3	*−8.2*	−1.1	−2.6	0.1	4.1
MC6	TC1	0.5	−4.9	*7.6*	1.3	−0.6	3.9
	TC2	−1.0	−8.0	*27.5*	1.3	−1.5	3.4
	TC3	0.7	−1.8	*51.5*	−0.0	−1.1	1.0
	TC4	−1.3	4.4	*62.3*	−2.0	3.0	−0.4

Note:
SC1 – Straight line flyer.
SC2 – Steady improvement in skills and tactics.
SC3 – High stick control.
SC4 – Slow reactors.
SC5 – Increasingly conservative tactics – relative to others.
SC6 – Increasingly poor tactics – relative to others.

component go from positive to negative over time, this component represents the steady improvement in performance over blocks displayed by virtually all subjects to varying degrees. That is, all subjects start out high in tactical ineffectiveness (i.e., measure component 1) in the early blocks and become better as they play more games. Similarly, on measure component 3 the subjects use conservative tactics during the early phases of training and become less conservative as they play more games.

Likewise, the second subject component indicates improvement in foe mine identification (i.e., measure component 4) and interval timing (i.e., measure component 5) as more games are played. The remaining five subject components correspond to only one measure component and show increasing distinctiveness through the blocks. The interpretation of these subject components is straightforward (i.e., the same as the corresponding measure component). For instance, the *first subject component* corresponds only to the measure component 2 (i.e., straight line flight in the Y-axis) and indicates an increasing distinctiveness in this behavior in the later stages of training (as indicated by the increasing size of the loading across the block components).

Three of the measure components (1, 3, and 4) have large loadings corresponding with both subject component 2 (the general improvement component) and with another separate subject component. If the rate of improvement on one (or more) of these measure components is different from the rest of the measure components then the subject's loading on the specific subject component would indicate this different rate of change. For example, a subject with positive loadings on both subject components 2 and 4 is one who, although showing steady improvement over blocks overall, does not improve with respect to mine identification as fast as with other aspects of the game.

What and when a subject learns and what level of performance is achieved depends, in part, on the subject's initial ability as assessed by the aiming screening task. Correlating the subjects' screening scores with their loadings on the subject components produces significant negative correlations with three of these components: the first ($r = -0.45$), the fourth ($r = -0.40$), and the fifth ($r = -0.56$). This indicates that those subjects who have high screening scores tend not to pursue a straight-line flying strategy, and do not have difficulty in foe mine identification or interval timing.

A cross-validation of the results: The Israeli sample

The Three-Mode PCA provided a description of different aspects of subjects' behavior across blocks, including different strategies and tactics used during the play of the game. It also pointed out individual differences in performance. Although these results gave an adequate description of the structure found in the data obtained in the control group, it is not clear how well this structure will generalize. To examine this issue the model was applied to a second data set. The data from the subjects who served as a control group at the Technion, Haifa, Israel, were submitted to the model derived from the Illinois control subjects (see Gopher et al. 1989, this volume). This analysis was particularly interesting because there were remarkable differences between the Israeli and the American control groups. While the learning curves obtained from the two groups were essentially similar in shape, the Israeli subjects performed at a consistently lower level than did the American subjects. By applying the model discussed in the previous section to the Israeli group, we obtained a description of both groups within a common measurement space. This, we hoped, would illuminate the nature of the difference between the two groups. [8]

[8] Given that the Technion subjects did not come from the same population as the Illinois subjects and because their performance was consistently worse than that of the Illinois subjects, this application of the model to the Technion sample is not a 'true' cross-validation.

Table 7
Subject types.

Subjects	SC1	SC2	SC3	SC4	SC5	SC6
Type 1: Steady improvement with high joy-stick manipulation, and little straight-line flight						
C1	−0.28	*0.23*	*0.46*	−0.37	0.27	*−0.67*
C19	−0.28	*0.48*	*0.57*	−0.38	0.15	*−0.45*
C20	−0.15	*0.36*	*0.75*	−0.34	0.12	*−0.40*
C33	−0.14	*0.67*	*0.49*	−0.12	0.18	*−0.50*
C10	−0.19	*0.48*	*0.74*	0.71	0.28	*−0.28*
*G7	−0.21	*0.49*	*0.74*	0.25	0.15	*−0.28*
C28	−0.17	*0.44*	*0.75*	0.04	0.26	*−0.39*
C5	−0.06	*0.44*	*0.62*	−0.33	−0.55	*−0.06*
C23	−0.29	*0.44*	*0.39*	−0.17	−0.71	*−0.21*
Type 2: Steady improvement with low ship movement and no straight-line flight						
C6	*−0.54*	*0.75*	−0.01	−0.15	−0.19	−0.30
C29	*−0.53*	*0.67*	0.05	−0.18	0.39	−0.28
C31	*−0.71*	*0.41*	−0.41	−0.27	−0.04	−0.30
C34	*−0.63*	*0.50*	−0.31	−0.10	0.11	−0.48
C35	*−0.50*	*0.71*	−0.16	0.42	0.02	0.21
*G19	*−0.48*	*0.64*	−0.25	0.43	0.33	0.01
*G2	*−0.54*	*0.52*	−0.30	0.27	0.38	−0.35
*G8	*−0.54*	*0.37*	−0.12	0.30	0.46	−0.51
Type 3: Low ship movement, no straight-line flight and fast mine handling						
C12	*−0.45*	0.05	*−0.26*	−0.39	*−0.40*	−0.65
C14	*−0.57*	0.24	*−0.27*	−0.33	*−0.22*	−0.63
C17	*−0.57*	0.45	*−0.42*	−0.18	*−0.50*	−0.09
C21	*−0.33*	0.33	*−0.46*	−0.11	*−0.75*	−0.04
C18	*−0.48*	0.25	*−0.23*	−0.20	*−0.72*	−0.30
Type 4: Conservative tactics and good management of resources. Steady improvement, even though they start off very poorly						
C3	0.03	*0.57*	0.35	0.31	*0.67*	0.02
*G14	0.06	*0.49*	0.51	0.56	*0.41*	0.10
*G6	−0.19	*0.34*	0.03	0.53	*0.69*	0.29
*G16	0.31	*0.30*	0.04	0.60	*0.62*	0.27
C36	0.37	*0.34*	0.11	0.04	*0.84*	0.19
C37	−0.35	*0.60*	0.17	0.19	*0.68*	0.06
C40	0.13	*0.44*	0.30	0.20	*0.70*	0.40
*G3	−0.34	*0.50*	0.24	0.24	*0.71*	−0.15
C38	−0.24	*0.59*	0.50	−0.02	*0.58*	0.00
Type 5: Slow reaction times and bad mine identifications						
C26	0.01	0.49	0.21	*0.77*	0.25	0.22
*G4	−0.01	0.37	−0.12	*0.73*	0.38	0.41
C27	0.04	0.20	−0.10	*0.97*	0.08	0.07
*G1	−0.13	0.20	−0.12	*0.89*	0.37	0.08
*G17	0.23	0.30	−0.05	*0.85*	0.36	0.10
*G11	0.31	0.29	−0.27	*0.67*	0.05	−0.54

[*Continued on following page*]

Table 7 (continued)

Subjects	SC1	SC2	SC3	SC4	SC5	SC6
Type 6: Increasingly poor tactics (with respect to other subjects)						
C7	−0.06	*0.46*	0.36	0.20	−0.06	*0.78*
C22	0.18	*0.33*	0.65	0.15	−0.07	*0.64*
*G15	−0.16	*0.07*	0.94	−0.05	−0.01	*0.29*
C8	0.05	*0.43*	−0.09	0.52	−0.01	*0.73*
*G18	−0.00	*0.58*	−0.20	0.37	−0.23	*0.65*
C16	−0.19	*0.61*	0.26	0.31	0.31	*0.58*
C32	0.20	*0.72*	−0.03	0.17	0.39	*0.51*
C39	−0.03	*0.70*	−0.06	0.36	0.42	*0.45*
C11	−0.10	*0.49*	0.41	0.60	0.01	*0.47*
*G9	0.03	*0.42*	0.24	0.77	−0.12	*0.40*
*G10	0.09	*0.31*	0.20	0.83	−0.36	*0.18*
*G13	0.39	*0.40*	−0.07	0.42	−0.42	*0.57*
Type 7: Straight-line fliers						
C2	*0.54*	*0.33*	*−0.37*	−0.27	0.08	−0.62
C15	*0.50*	*0.30*	*−0.38*	−0.18	0.31	−0.62
*G5	*0.52*	*0.38*	*−0.40*	−0.28	0.33	−0.48
*G12	*0.47*	*0.43*	*−0.43*	0.03	0.31	−0.56
C4	*0.29*	*0.52*	*−0.48*	0.11	0.62	0.14
C25	*0.59*	*0.28*	*−0.55*	−0.01	0.42	−0.31
C30	*0.69*	*0.33*	*−0.54*	0.05	0.35	−0.06
C9	*0.76*	*0.31*	*−0.52*	0.21	−0.07	−0.03
C24	*0.69*	*0.53*	*−0.47*	0.00	0.11	−0.02
C13	*0.66*	*0.47*	*−0.49*	−0.14	−0.16	−0.22

Note:

Subjects: C indicates subjects in the Illinois control group. G indicates subjects in the Israeli control group run by Gopher and colleagues. The Israeli subjects are marked by a *, for prompter identification.

SC: The acronym SC, followed by a number, indicates a specific subject component.

Nineteen students at The Technion, Haifa, Israel participated as control subjects in a study conducted by Gopher et al. (1989, this volume). The same procedures used for the Illinois control subjects, reported above, were employed with the Israeli group. The Technion subjects performed the aiming screening task more poorly, on average, than did the Illinois subjects. The Israelis had a mean screening score of 982.1 and a standard deviation of 89.2 (maximum score of 1,160).

The loading for the measure and block components and core matrix were fixed to those found for the Illinois sample. Only subject loadings were estimated. This was a simple least squares multiple regression problem. Note that in eq. (1), once the values of h, e, and c are fixed, the resulting equation simplifies to a standard linear regression equation. This constrained model explained 36% of the variance in the Technion subjects as compared to the 60% in the original sample.

Cluster analysis

In order to facilitate the comparison of subjects, since it is easier to think of types of subjects rather than weighted components, a cluster analysis was performed on the normalized subject loadings from both groups. The results of an equal variance maximum likelihood clustering method (SAS Institute 1985) are reported here. [9] Seven clusters were found to yield a coherent structure (and were supported by a log-likelihood ratio test and by their interpretability). Loadings for these clusters are presented in table 7.

Subject type 1 is a cluster of 9 subjects (8 from Illinois and 1 from the Technion) with positive weights on subject component 2 (SC2) and 3 (SC3) and negative weights on subject component 6 (SC6). These are subjects who showed steady improvement in game play and whose strategy involved high stick manipulation and little straight-line flight (i.e., a circling of the fortress while remaining on the screen).

Subject type 2 (5 subjects from Illinois and 3 from Technion) also exhibited steady improvement (positive loadings on SC2). The flight strategy for these subjects involved very little ship movement on the Y axis (negative loadings on SC1), and they also tended to move the ship very slowly (negative loadings on SC3).

Subject type 3 (5 subjects from Illinois) are those subjects who showed even less ship movement then the second type (negative on both SC1 and SC3. These subjects also fired the highest number of shots (and since they were moving very slowly they were very effective in hitting the fortress) and were quick to energize friendly mines or to destroy foe mines (i.e., negative loadings on SC5).

Subject type 4 (5 subjects from Illinois and 4 from Technion) are those subjects who were conservative in their tactics (i.e., few shots, long time to destroy the fortress or mines) and took advantage of the bonus opportunities (positive on SC5). These subjects started out much slower than the others (e.g., took longer to destroy the fortress) but improved at a faster rate so that by the end of the training they were performing about average (positive on SC2).

Subject type 5 (2 subjects from Illinois and 4 from Technion) is similar to type 4 (i.e., few shots, many missed, long time to destroy the fortress, etc.), although they did not start off as badly. They are distinguished by slow reaction times and bad identifications of mines (positive on SC4).

Subject type 6 (7 subjects from Illinois and 5 from Technion) are subjects who consistently got shot by the fortress, hit by a mine, and missed their target more frequently than the others (positive on SC6).

Subject type 7 (8 subjects from Illinois and 2 from Technion) are subjects who controlled the flight of their ship very little, moving consistently along the Y-axis, and jumped into hyperspace frequently (positive on SC1 and negative on SC3). As a result, they moved into the fortress' line-of-fire frequently.

Note that the Cluster Analysis presented here provides a more detailed subject classification than the strategy analysis reported in the first section of this paper. However, there is a clear correlation between the two analyses. For example, type 3

[9] A comparable clustering solution was also obtained using the complete link method.

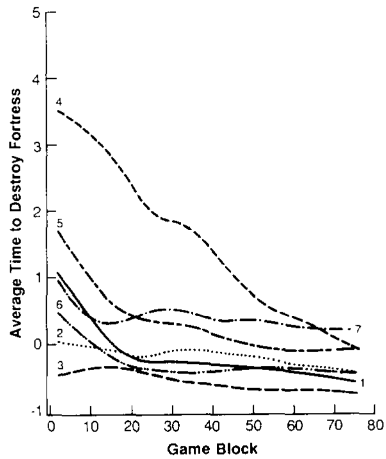

Fig. 8. Learning curves for subjects belonging to different subject types. The average time to destroy the fortress (in standard units) is plotted on the ordinate. The game block is plotted on the abscissa. Note that a decrement in the score represents an improvement of performance.

subjects are mostly those who used a very slow circling strategy. Also, subjects who used a straight-line flight pattern are all grouped in subject type 7. Table 3 reports, for each of the subjects, the corresponding subject type as identified by the Cluster Analysis.

The comparison between the strategy classification based on the raw data and the results of this Cluster Analysis provides an indirect way of validating the results of the latter. In addition, Cluster Analysis provides a broader way of defining and classifying strategies, because it takes into account other aspects of the subject's behavior that covary with his ship maneuvering strategies.

Further understanding of these subject types can be obtained by examining their performance on the individual measures over blocks. For eight of the measures ($NSHDMG_1$, SHOTEFF, MEAN RT, STDDEVRT, AVKTFR, MINTERV, STDEVINT, and PRCNBON), although the level of the various subject types differ, the rate of change across blocks does not. These are aspects of performance in which all subjects seem to improve at a uniform rate. For the other measures there are clear differences between subject types. For example, in fig. 8 is plotted (using a spline fitting procedure) the average time to destroy the fortress for each block. For most of the subject types much of the improvement occurs within the first three sessions (i.e., by game number 20) but for subject type 4 there is continual improvement on this measure throughout all the blocks. Fig. 9 depicts the frequency of hitting the fortress in a game. It can be clearly seen that subject type 3 consistently outperforms the others on this measure and that subject type 1 shows the greatest rate of improvement.

The performance of the Technion subjects, as measured by total score, was lower than for the Illinois subjects early in training. Furthermore, the Technion subjects improved at a lower rate across blocks. The average score for the Technion subjects

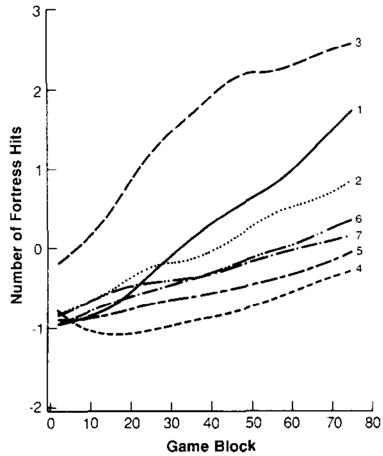

Fig. 9. Learning curves for subjects belonging to different subject types. The number of fortress hits (in standard units) is plotted on the ordinate. The game block is plotted on the abscissa.

during the first stage (i.e., the first 5 blocks) was -1349.9 ($sd = 558.9$) and for the Illinois subjects was -1040.7 ($sd = 859.6$). During the last stage (i.e., the last 28 blocks) the Technion subjects averaged 1665.4 ($sd = 888.2$) while the Illinois subjects averaged 2335.3 ($sd = 1004.5$). The performance, based on total score, of each of the subject types is shown in fig. 10. Type 3 subjects consistently performed the best, while types 1, 2, and 7 were in the middle, and types 4, 5, and 6 performed the worst. The

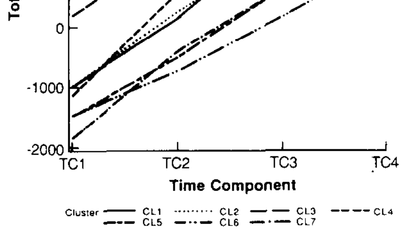

Fig. 10. Learning curves for subjects belonging to different subject types. The total score is plotted on the ordinate. The time components are plotted on the abscissa.

differences between the Illinois and Technion subjects are also evident in these results. An examination of the distribution of subjects into the various types shows that 13 of the Technion subjects (more than may be expected by chance) occur in types 4, 5, and 6 (the three worst performing groups) and none in subject type 3 (the best performing group).

Discussion

Forty students at the University of Illinois were instructed to practice the Space Fortress game for 10 successive one-hour training sessions. This 'control' group was given the opportunity to develop its skill without benefit of explicit training regimes. This, of course, is the common approach to the use of such skill trainers. The device provides a setting in which the subject can practice the task and it is assumed that the exposure, and the continuing practice, will help the subject in acquiring, perfecting and maintaining the skill. As the data show, subjects are indeed able to improve in such a regime. An examination of the data from all 40 subjects reveals that, without exception, all subjects were able to show enormous progress. Upon first exposure the task appeared dauntingly difficult. Subjects invariably failed to survive for more than a few seconds. Yet, with practice, they began to develop competence. Their total score continued to improve and it appeared that even after the 10 hours of practice their performance had not reached an asymptote.

It is important to examine in detail the nature of the progress the subjects make as they proceed, on their own, to meet the challenge of the task. One of the most important conclusions from our examination of the control groups is that there are systematic individual differences between subjects in the manner in which they approach this challenge. The total score, while a necessary metric for across-projects comparisons, obscures much that is important in the evaluation of the training. Subjects' strategies and the relative priority which they assign to different aspects of the task have much to do in determining the shape of the learning curve.

The realization that subjects do vary in this manner creates the need for a descriptive structure which can reduce to manageable size the massive data base acquired in such studies. With a computer driven task such as Space Fortress it is rather easy to acquire a very large number of measures that capture the instant by instant variation in performance. It became evident as we proceeded to compare the results of the various projects described here that an economical, yet multidimensional, metric that would capture the variance in performance would be beneficial. It was for this purpose that we subjected the data from the control group to the Three-Mode PCA. The results were not available for use across the entire project's data base. Yet, as a general tool for the analysis of such complex data bases this approach seems to hold much promise.

There are two different strategies for using Three-Mode PCA in this context. One approach is to apply the analysis afresh to each data set, obtaining in each case a description unique to the analyzed data. In this case, comparisons across groups require an assessment of the similarity between multidimensional structures. The other approach, and the one we preferred, utilizes the description yielded by an analysis of the data obtained from a reference group as a baseline against which to compare the data obtained from other groups. Rather than obtaining a new structure for each group, we try to locate all compared groups within the space defined by the reference group.

In the present study, the Three-Mode PCA of the Illinois control group provided us with a space in which 40 subjects could be replaced by 6 subject components, the 75 game blocks over the 10 hours of testing could be reduced to 4 time epochs, and the many measures of the subjects' performance reduced to 6 measure components. A space with $6 \times 6 \times 4$ dimensions cannot be considered 'simple'. Yet, it is simpler than a space of $40 \times 75 \times 21$ dimensions. The utility of this simplifying strategy is emphasized by our ability to examine the intriguing differences that were observed between the Israeli subjects and the 40 subjects used in this project.

The subjects run at the Technion performed consistently worse than subjects tested in all other Learning Strategies projects. Technion subjects had lower scores on the aiming screening task. More candidates were rejected before they were even admitted to the project because their screening score fell below criterion. The learning curves of the Technion subjects were invariably lower than the curves of other subjects. These data are remarkable because the Israeli subjects were mostly students at the Technion, a highly selective engineering college. By and large, all served in the Israeli army prior to entering college and were often exposed to complex electronic equipment and complex tasks. Thus, it is useful to try and assess the specific nature of this difference. The analysis allowed by the Three-Mode PCA suggested that the difference was localized rather specifically to those aspects of subjects' behavior that were related to the speed of performance. The Israeli subjects were concentrated specifically in those segments of the space that were characterized by slow reactions to mines and to the fortress, and generally conservative tactics, and by limited ability to control the ship. It is also interesting to note that no Israeli subjects were included in the best performing group (type 3), which was characterized by a high degree of ship control and by efficient shooting performance.

These facts, in themselves, do not explain the results because it remains to be determined what it was about the Israeli subjects that located them in this sub-domain of the space. However, the results of the Three-Mode PCA can be used to direct the search for the cause of the differences better than could the total score alone, or an attempt to use all the data variables without reduction. Thus, this analysis illustrates the utility of the Three-Mode PCA in reducing the dimensionality of very complex data bases.

References

Cliff, N., 1968. 'Idealized Individual' interpretation of individual differences in multidimensional scaling. Psychometrika 33, 225–232.

Donchin, E., 1989. The Learning Strategies project: Introductory remarks. Acta Psychologica 71, 1–15 (this volume).

Fabiani, M., J. Buckley, G. Gratton, M.G.H. Coles, E. Donchin and R. Logie, 1989. The training of complex task performance. Acta Psychologica 71, 259–299 (this volume).

Gopher, D., M. Weil and D. Siegel, 1989. Practice under changing priorities: An approach to the training of complex skills. Acta Psychologica 71, 147–177 (this volume).

Kaiser, H.F., (1958). The varimax criterion for analytic rotation in factor analysis. Psychometrika 23, 187–200.

Kroonenberg, P.M., 1983. Three-mode principal components analysis: Theory and applications. Leiden: DSWO Press.

Kroonenberg, P.M., 1984. 'Three-mode principal components analysis: Illustrated with an example from attachment theory'. In: H.G. Law, C.W. Snyder, Jr., J.A. Hattie and R.P. McDonald (eds.), Research methods for multimode data analysis. New York: Praeger. pp. 64–103.

Kroonenberg, P.M. and P. Brounwer, 1985. Programmers guide to TUCALS (version 4.0). (WEP Reeks WR-85-07-RP). Leiden: Dept. of Education, University of Leiden.

Kroonenberg, P.M. and J. de Leeuw, 1980. Principal components analysis of three-mode data by means of alternating least squares algorithms. Psychometrika 45, 69–97.

Mané, A.M., 1985. Adaptive and part-whole training in the acquisition of a complex perceptual-motor skill. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, USA.

Mané, A.M. and E. Donchin, 1989. The Space Fortress game. Acta Psychologica 71, 17–22 (this volume).

Mané, A.M., J.A. Adams and E. Donchin, 1989. Adaptive and part–whole training in the acquisition of a complex perceptual-motor skill. Acta Psychologica 71, 179–196 (this volume).

Mané, A.M., M.G.H. Coles, D. Karis, D. Strayer and E. Donchin, 1984. The design and use of subtasks in part training and their relationship to the whole task. Proceedings of the 28th annual conference on manual control.

SAS Institute, 1985. SAS user's guide: Statistics (1985 ed.). Cary, NC: SAS Institute.

Tucker, L.R., 1966. Some mathematical notes on three-mode factor analysis. Psychometrika 31, 279–311.

Tucker, L.R. and C.T. Finkbeiner, 1981. Transformation of factors by artificial personal probability functions. (Research report RR-81-58.) Princeton, NJ: Educational Testing Service.

Tucker, L.R. and S. Messick, 1963. An individual difference model for multidimensional scaling. Psychometrika 28, 333–367.