ERGONOMICS, 1997, VOL. 40, NO. 3, 362–380

Information complexity—mental workload and performance in combat aircraft[1]

E. SVENSSON,† M. ANGELBORG-THANDERZ,† L. SJÖBERG‡ and S. OLSSON†

†National Defence Research Establishment, Linköping, Sweden
‡Center for Risk Research, Stockholm School of Economics, Sweden

Keywords: Mental workload; Pilot performance; Information processing; Pilot decision making; Situational awareness.

The purpose of the present study was to analyse the effects of information complexity on Pilot Mental WorkLoad (PMWL) and Pilot Performance (PP), and to analyse the structure of PMWL. Eighteen pilots performed 72 simulated low level–high speed emissions. The complexity of the Head Down Display (HDD) information varied as a function of the tactical situation. Flight data were recorded continuously. The pilots' eye movements were video taped and psychophysiological activation data, Heart Rate (*HR*), were obtained. The pilots rated PMWL according to the psychological content of three scales (Bedford Rating Scale, Subjective Workload Assessment Technique, NASA-Task Load indeX) and answered a questionnaire tapping aspects of performance, information load, motivation and mood. It was found that even a moderate complexity of information interfered with the flight task. Altitude and variation in altitude were increased and corrections of altitude errors were delayed, when complexity increased. Changes in information load reached its maximum influence on flight performance ($r = 0{\cdot}60$, $p < 0{\cdot}001$) after a time delay of 20 to 40 seconds. Performance of the flight task correlated positively ($r = 0{\cdot}59$, $p < 0{\cdot}001$) with the performance of the information handling task (Tactical Situation Awareness, TSA). Durations and frequencies of eye fixations Head Up (HU) versus Head Down (HD) changed as a function of information load. A structural equation model implied that PMWL was affected by mission complexity and that PMWL affected objective and subjective aspects of flight performance and information handling. Heart rate (sortie means) correlated positively with PMWL ($r = 0{\cdot}34$, $p < 0{\cdot}05$) and perceived complexity of mission ($r = 0{\cdot}37$, $p < 0{\cdot}01$). Heart rate (running means) covaried with variations in information complexity for those pilots who performed well. From spectral analyses of cardiac interval times it was found that the amplitude of the 0·10 Hz component tended to decrease during high as compared to low levels of information load.

1. Introduction

1.1. *Decisions and complexity of information*

Decisions are affected by uncertainty, ambiguity and limited human capacity. It is hardly possible for anyone to be observant of everything at the same time. Even if the military pilot is rigorously selected and trained, the corresponding demands on him or her to react rapidly and accurately in a synthetic and 'hyperdynamic' environment are extreme. For this reason, it is important to take into account what is known about limitations of human information processing.

When too many alternatives with too many attributes compete for attention, the decision maker will be mentally overloaded. Mental overload often has its origin in

[1]The study has been performed by order of and in close cooperation with the Swedish Air Force.

cognitive limitations. Miller (1956, cf. Baddeley 1994) found that humans cannot discriminate between more than half a dozen one-dimensional entities; nor can they handle more objects in their short term memory or control more content in their attention.

Moray (1986) cited research where attempts were made to decide what is the optimal number of elements in decision making in uncertain situations. Normally it is not an advantage to have more than seven elements. The same conclusions are valid for judgement and estimation. The number seven appears also in applied research on pilot performance. Eustase (1977) found that when more than seven simultaneous threats were presented to the pilot, his outcome, in terms of number of omissions and errors, radically deteriorated. Way *et al.* (1987) stressed the importance of reducing display clutter and making critical information more evident.

According to Moray (1986) the size of dynamic working memory is about three items for random and meaningless time series. However, the limitations of the working memory are not strictly fixed; even if the number of units is limited, the content of every unit seems to be almost unlimited! Accordingly, if there is a need to handle abundant information, it should be structured in connected chunks instead of discrete bits (Chase and Ericsson 1981, Simon 1990).

If operators must choose between two simple alternatives, they can make a maximum of two decisions/second (2 bits/second). However, if they can use their knowledge of more elaborate information, e.g. by means of language, they can handle a much larger flow of information, 8–10 bits/second (Wickens 1987). The expression *decision complexity advantage* means that performance will be much better if one arranges the information in such a way that the operator can make fewer complex decisions rather than a large number of simpler ones.

Many people however, feel safer with more information, even if they do not use it. There is an *illusion of knowledge* (van Raaij 1988). The quality of the decisions decreases with increasing amount of information beyond the optimum, at the same time as the decision maker's illusion increases.

The time factor is in itself a stress factor and, of course, important in the analysis of the pilot's judgements and decisions in a rapidly changing environment. Generally, psychological stress can induce tunnel vision and a more primitive motor behaviour (Easterbrook 1959). It can also cause an emphasis on negative and threatening information (Svenson and Edland 1987). On the other hand it is not certain that availability of more time will lead to better decisions (Svenson and Benson 1991).

Those pilots who can best integrate 'performance critical' information have a Position Of Advantage[2] (POA). This ability forms the core of the concept Situational Awareness (SA) for which the skill aspect is of importance. Even if the theoretical anchorage of SA is still weak, it can be assumed that it calls for preattentive and automatic information processing and an efficient memory function (cf. Gilson 1995).

Klein (1993), in his naturalistic decision making models, regards the pilot's ability to recognize and assess the situation (situation assessment) as the crucial factor of the decision making process. For the skilled pilot almost every situation is associated with a 'best alternative'.

The limitations on human information processing have been known for a long time. In spite of this, designers are apparently tempted by the possibilities, created by modern computer technology, to include increasingly complex and numerous options (modes) and

[2]Jane's Aerospace Dictionary 1986.

displays in their systems. The result is that the human operator is faced with a very complex task, here illustrated by combat aircraft pilots.

1.2. *Pilot Mental Work Load (PMWL), its rationale and measurement*

Pilots in modern military flight systems have to process a considerable amount of complex information, much more comprehensive than in older systems. The information and decision making processes have become more and more demanding and the risk of mental overload has increased (Angelborg-Thanderz 1990). Procedures for measuring PMWL have become an indispensable prerequisite for analysing specific missions and evaluating the need for decision support systems (data fusion, artificial intelligence and expert systems).

The development of combat aircraft calls for measures not only of Pilot Performance (PP) but also of PMWL, and a lack of standardized and accepted measures makes it difficult to compare different systems design and their operational requirements. The relation between performance and workload represents a measure of human efficiency. The assessment techniques presented have all their advantages and disadvantages. Unfortunately, none of them has a transparent theoretical basis.

Subjective ratings seem to be the most used technique, sometimes used in combination with a physiological measure (especially heart rate) (Roscoe 1987, Roscoe and Ellis, 1990, Eggemeier and Wilson 1991, Wilson and Eggemeier 1991). Casali and Wierwille (1983) recommend subjective techniques, because they are cheap, easily administrated and adaptable to different situations. Modified versions of the Cooper-Harper Handling Characteristic Scale (Cooper and Harper 1969, Wierwille and Casali 1983, Wierwille and Connor 1983) have frequently been used in the aircraft industry in measuring PMWL.

The BedFord Rating Scale (BFRS) is a decision tree scale derived from the Cooper-Harper Scale. According to Lysaght *et al.* (1989) 'The technique obtains subjective judgements about workload based on ability to complete tasks and the amount of spare capacity available' (p. 88).

The Subjective Workload Assessment Technique (SWAT) (Reid and Nygren 1988) and the NASA Task Load IndeX (TASA-TLX) (Hart and Staveland 1988) exemplify frequently-used multidimensional rating scales. In SWAT the aspects of time load, mental effort load and stress load are merged into a workload index by means of conjoint measurement. NASA-TLX measures the aspects of mental, physical and temporal demands, as well as performance, effort and frustration levels. The aspects are differentially weighted and merged into a workload index.

The subjective measures of PMWL are, more or less, static (i.e. they reflect mean levels over periods of minutes or more). This is a serious drawback in analyses of (hyper)dynamic situations (e.g. flight missions) and when dynamic subjective measures are called for (cf. Jensen 1994).

It is sometimes claimed that the reliability and validity of subjective ratings of mental workload (and performance) are insufficient (Muckler and Seven 1992). All mental processes are not introspectively available and, accordingly, subjective measures may underestimate workload. Retrospective ratings can be affected by memory deficiencies (Gopher and Donchin 1986, O'Donnell and Eggemeier 1986) and overestimation can also occur due to context effects in the application of subjective measures.

Subjective measures have, however, turned out to be extremely useful in many contexts. Doubts about their validity, while sometimes justified, should not be exaggerated. Even if the precision of any single rating is modest, data may still be sufficiently rich in information to be useful. Reliabilities of the mood scales used in the present study have been found to range between 0·65 and 0·95 (Sjöberg *et al.* 1979). In comparison, the covariations between different psychophysiological measures are often low or insignificant and stimuli of different kinds elicit an unspecific psychophysiological reaction.

Objective measures of performance sometimes correlate strongly with perceived workload or effort, sometimes not. However, this dissociation cannot be taken as an indication of measurement problems when it comes to workload *per se*. Performance and workload measures are sensitive to different task factors (Wickens and Yeh 1983, Gopher and Donchin 1986).

In the present study we have used subjective assessments and estimates of psychophysiological activation as indicators of PMWL. We have used and compared the psychological content of three PMWL-scales: (a) BFRS (Roscoe 1987, Roscoe and Ellis 1990), (b) SWAT (Reid and Nygren 1988) and (c) NASA-TLX (Hart and Staveland 1988). Heart rate (sortie means, running means and variance) has been used as an indicator of the subjects' psychophysiological activation during the missions.

The purpose of the present study was to: (a) analyse the pilot's ability to perceive and use synthetic information, (b) by means of linear causal modelling study the effects of information complexity on PMWL and PP, and (c) by means of different indices analyse the structure of PMWL.

2. Method

Since the seventies we have used the Swedish Air Force flight training simulators as a research platform in analyses of pilot behaviour and development of models of PP (Thanderz 1973, 1982, Angelborg-Thanderz 1989, 1990). In recent years, interest in simulation as an important research tool has grown in studies of human performance in realistic and complex situations. In the present study, the information of the Tactical Situation Display (TSD) of two JA37-simulators was modified with respect to the symbols complexity.

Eighteen active fighter pilots participated and they performed in all 72 sorties. The pilots' mean age was 31 years and their mean time on military aircraft was 870 hours. The range was large with respect to age (from 22 to 51 years) as well as flight experience (from 100 to 2700 hours).

The pilots were given two main tasks during low level–high speed flight; to fly and to handle tactical information. The *flight task* comprised navigations over a sometimes hilly landscape with a number of waypoints and at a specified altitude and speed. The *information handling task* comprised perception and control of own and enemy air defence areas and own and enemy aircraft presented on the TSD (i.e. presentation of the tactical situation). The pilots had to report changes continuously (verbally via 'open line') with respect to the number and type of 'objects' presented. Their reports were recorded, adapted manually and saved on disks.

Three variants of missions with tactical information on TSD were performed (the 2nd, 3rd and 4th run). These runs differed with respect to level of (number of own and enemy defence areas and aircraft) and variability of (rate of change of defence areas and aircraft) situation complexity. The mean level and variation in information load was highest in the second run and lowest in the third run. Topographically the third run was the most complex and the fourth the least. The second run combined a high and varying information load with a moderate topographical complexity. Each mission lasted about 30 minutes. The sequences were ordered according to a latin square, repeated measurement design. The first run or sortie had a training goal. No TSD-information was presented during this run and the flight performance represents a reference unaffected by information handling.

The prescribed altitude was set to 200 m and the prescribed speed to 750 kph. The pilots were ordered to keep the prescribed altitude and speed without using technical aids. The scenarios of the runs were based upon realistic situations.

A short briefing was given to the pilots before each run. A profile of the navigations was presented and the pilots were instructed how to report changes of the tactical situation presented on TSD (i.e. the number of, or the change of number of, own and enemy defence areas and aircraft). They were instructed to make their own priority with respect to the flight- and information-handling tasks.

The pilots answered check-lists before and after each sortie. Before the sorties they rated their motivation, expected performance and mood in terms of the following dimensions: hedonic tone, activation, tension and control (Sjöberg *et al.* 1979, Svensson *et al.* 1988, 1993).

At four times during the sorties the pilots rated the three aspects of SWAT: time stress, mental effort and psychological stress. The ratings were made on a seven-point Likert-scale instead of the conventional three-point scale. The three aspects were equally weighted to an index of momentary PMWL called 'SWAT'. The pilots reported their ratings verbally after periods of varying length (with a range of 6 to 10 minutes) and of varying information load (a range of mean number of 'threats' from 3 to 10).

All other items were responded to after the sortie in the following order: 60 assorted items, sortie mood, and six items of NASA-TLX, BRFS and final mood. The subjects answered all these items under supervision; back checking did not occur.

The questionnaire is tapping aspects of difficulty, performance, motivation, control, vigilance, mental capacity, mental and physical effort, concentration, information load, priority to tasks, interference between tasks, availability and complexity of information (in all 60 items). By means of iterative principal axis factor analysis the number of items was reduced to ten indices. The reliability of the indices was tested by means of Cronbach's alpha. The six aspects of NASA-TLX as well as the items of the questionnaire employed a seven-point response format. The items of NASA-TLX were equally weighted. The mean of the items were called 'NASA-TLX'.

Psychophysiological activation was registered in terms of heart rate (HR). The pilots' eye movements were video taped (Philips CCD-camera LDH 0460/00). The durations and frequencies of eye fixations HU and HD were recorded manually from the video tape.

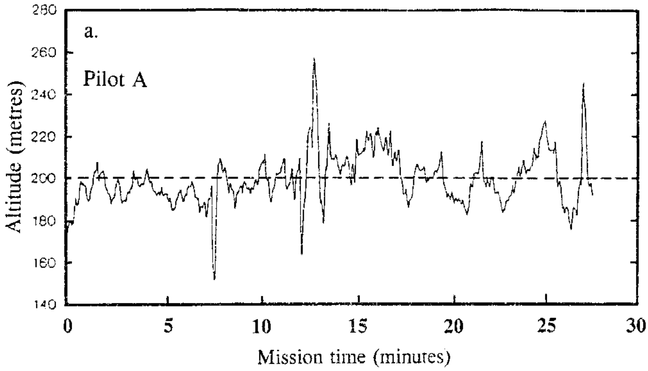
Primarily, correlational statistics (simple and multiple regression, factor analysis, second generation multivariate statistics and Fourier analysis (FFT)) were used. Linear causal model analyses were performed by means of LISREL VI (Jöreskog and Sörbom 1984). This procedure makes a statistical test of the validity of different causal flow models possible.

3. Results

3.1. *Flight performance and information handling*

Figure 1 presents flight performance (non-filtered data) in terms of altitude as a function of mission time for two pilots during the second run. Figure 1(a) represents the performance of 'an expert' pilot (pilot A) and figure 1(b) the performance of a less skillful pilot (pilot B). Pilot A represents the upper end of a rank order of the pilots with respect to flight performance and information handling. Pilot B represents the lower end of the same rank order.

It can be seen from the figures that the magnitude and variation in altitude were smaller for pilot A than for pilot B. Pilot B flew above the prescribed altitude (200 metres) most of



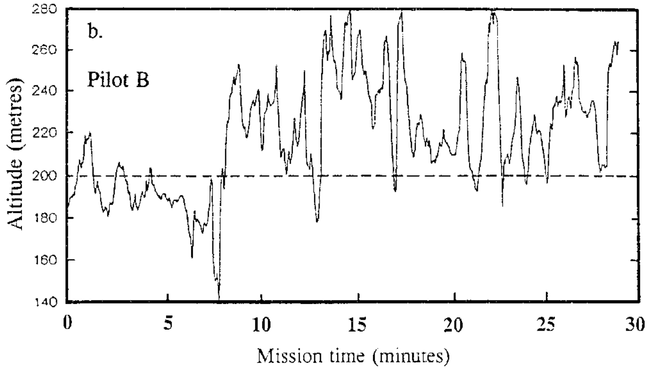


Figure 1. Altitude as a function of mission time for (a) pilot A and (b) pilot B. Pilot A represents a skilled pilot and pilot B a less skilled one.

the time. Furthermore, the variation in flight performance seems to increase as a function of mission time for both pilots.

When comparing the variation in information load with the flight performance in figure 1 we surmised that the pilots' flight performance covaried with the information load on TSD; the flight level increased when the number of objects on TSD increased.

Figure 2 presents the flight performance in terms of altitude as a function of mission time for the 18 pilots during the second run. The individual curves have been filtered and summarised to a single curve using distance weighted least squares regression (DWLS). The variation in information load on TSD is presented below the curve.

As can be seen from Figure 2 there was a covariation between altitude and number of objects presented on TSD (no systematic relationship was found between the density distribution of number of objects on TSD and flight performance). From analyses of auto-correlation and cross-correlation functions it was found that the covariation reached its maximum ($r = 0{\cdot}60$, $p < 0{\cdot}001$) when the time delay of the flight performance curve was reduced 20 to 40 seconds. Exclusion of the common trends of the curves reduced the covariation ($r = 0{\cdot}38$, $p < 0{\cdot}001$). This means that changes in information load reached its maximum effect on flight performance after 20–40 seconds. Increases in information load resulted in altitude increments and decreases in information load resulted in altitude decrements. The time delay supports the causal relationship between information load and flight performance.

Figure 3 presents altitude as a function of the number of objects presented on TSD. The curve summarises the flight performance of the 18 pilots during the second run. The

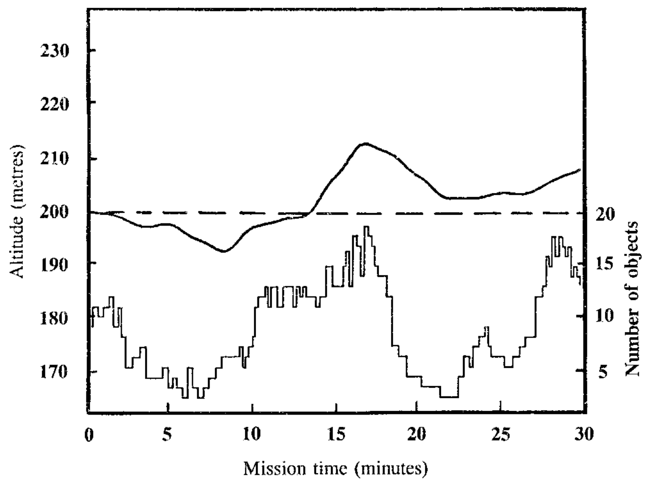


Figure 2. Altitude (the curve) and information load (the histogram) as a function of mission time for 18 pilots during the second run. The data have been filtered and summarized to a single curve by means of Distance Weighted Least Squares (DWLS) regression analysis.

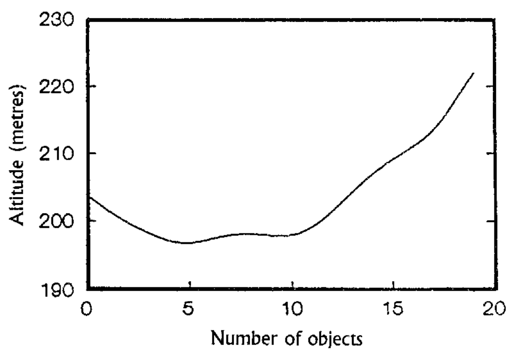


Figure 3. Altitude as a function of number of objects on TSD for the 18 pilots during the second run. The curves have been smoothed by means of DWLS.

second run has the largest range of 'threats' (2–18) and the landscape was flat as compared to the third run, and hilly as compared to the fourth.

The curve shows that the pilots flew on prescribed altitude (200 metres) up to a complexity of about 10 objects on TSD. With a further increase in complexity, the pilots climbed about 2 m for every new object presented. The correlation between altitude and complexity was insignificant ($r = -0{\cdot}12$, $p = 0{\cdot}12$) for the segment 2 to 10 objects. The corresponding correlation for the segment 11–18 objects was significant ($r = 0{\cdot}50$, $p < 0{\cdot}001$). The absolute deviation from prescribed altitude followed the same pattern. Thus, the flight performance was constant up to a certain complexity level on TSD and deteriorated, when the complexity increased further.

Figure 4 presents the flight performance in terms of altitude of pilot A and pilot B as a function of number of objects presented on TSD. The figure shows that pilot B flew on a higher altitude and climbed earlier than pilot A, when the information load increased. We found from analyses of flight precision (deviations from prescribed altitude) that the performance of pilot A was quite constant and independent of the variation in information

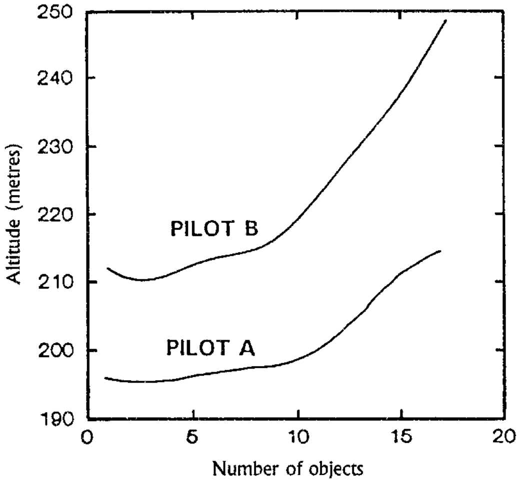


Figure 4. Altitude as a function of number of objects on TSD for pilot A and pilot B. The curves have been smoothed by means of DWLS.

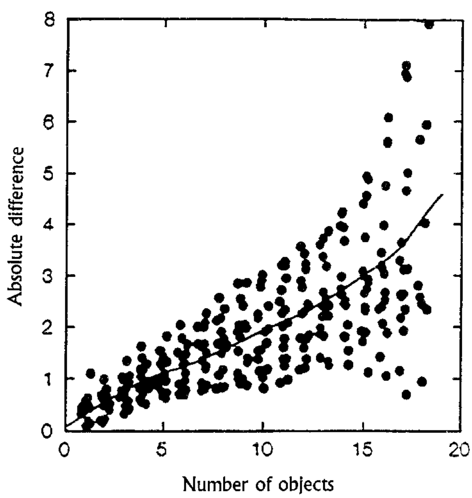


Figure 5. Absolute difference between the number of reported and presented objects as a function of number of objects. The circles represent individual values. The curve has been smoothed by means of DWLS.

load. Pilot B, on the other hand, had a marked deterioration in the performance when the information load exceeded about eight objects.

It can be concluded from the figures presented that the pilots flight performance deteriorated as a function of the information load emanating from TSD. The results presented were based upon data from the second run, but the conclusions were the same for the third and fourth run, even if the effects were less conspicuous.

Figure 5 presents the absolute difference between the number of reported objects and the number of presented objects as a function of the number of objects for the second run. The curve is derived by means of a distance weighted least squares regression analysis. As noted above, this run has the largest range of 'threats' (2–18).

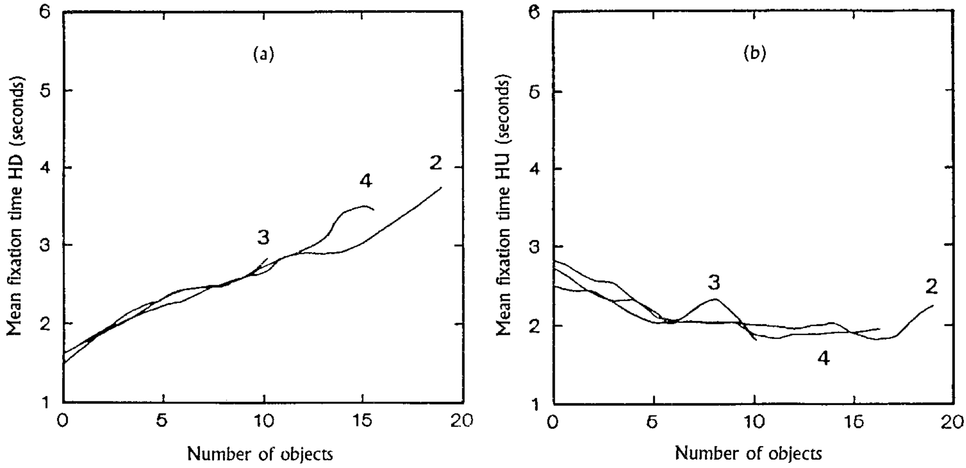


Figure 6. Mean fixation times HD (a) and HU (b) as a function of number of objects on TSD for the second, third and fourth runs. The curves have been smoothed by means of DWLS.

The figure shows that the absolute error increased as a function of information complexity. The error increased by one object when the complexity was increased by five objects, on average. Furthermore, the interindividual variance increased as a function of the complexity. Further analyses revealed that over- and underestimation cancelled each other out up to about 10 objects. When the number of objects increased further, the underestimations began to dominate and the ability of the pilots to detect changes in the tactical situation deteriorated drastically.

Comparisons between pilots A and B revealed that pilot B began to underestimate the number of objects, i.e. to miss critical objects or threats, at a complexity of about six objects. The corresponding value for pilot A was 15.

3.2. *Eye movements and fixations*

The flight task as well as the information-handling task are extremely dependent on vision. To investigate the pilots' strategies and mental workload when solving the two tasks, we registered the durations and frequencies of their eye fixations HU and HD.

Comparing the different types of runs, we found that the frequencies of shorter fixation times HU *and* the frequencies of longer fixation times HD increased as a function of the average information load on TSD. Thus, the conditions for flying low level–high speed with high precision deteriorated when the information load HD increased.

Figure 6(a) and (b) present the mean fixation times HD and HU as a function of the number of objects presented during runs 2, 3, and 4. Figure 6(a) shows that the fixation time HD increased approximately linearly as a function of the number of objects presented during the three runs (with a range r of 0·52–0·69, $p < 0{\cdot}001$). Figure 6(b) shows that the fixation time HU decreased up to about six objects (with a range r from $-0{\cdot}22$ to $-0{\cdot}29$, $p < 0{\cdot}01$). The fixation time HU was approximately constant (about 2 s) from six to 18 objects (with a range r of 0·08–0·11, n.s.). There seems to be an inverse relationship between the fixation times HU and HD up to about six objects.

We have considered fixation times HD with a duration longer than 4 s to be critical with a flight safety point of view (Goebel 1983). The percentages of 'critical fixation times' were 2·01, 19·94, 16·11 and 21·78 for the first, second, third and fourth runs respectively. These percentages were in concordance with the mean number of 'threats'

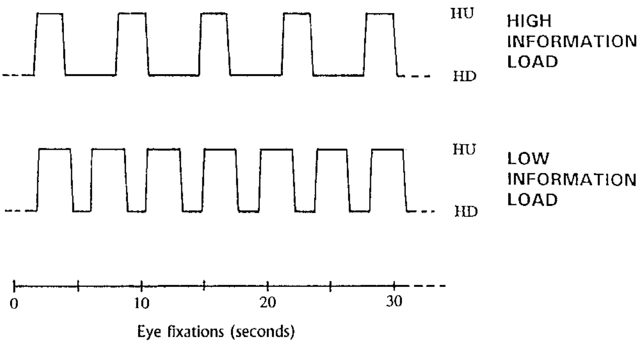


Figure 7. Eye fixations HU and HD over time, under high and low information load, respectively. The fixation times are based upon extrapolations of the regression lines in figure 6. High information load represents 20 objects on TSD and low information load no objects on TSD.

presented on TSD during the second, third and fourth run (7·53, 4·96 and 7·62, respectively).

Figure 7 illustrates how the average pilot, in a statistical sense, changes eye fixation HU-HD over time under high and low information load, respectively. The fixation times are based upon extrapolations of the regression lines in figure 6. The inspection times HU were shorter and their frequency was lower under high, as compared to low, information load. The durations and frequencies of critical eye fixations HD covaried with the PMWL measures (e.g. $r_{\text{freq. critical fixations HD — BFRS}} = 0{\cdot}51$, $p < 0{\cdot}01$).

3.3. *Performance and mental workload (PMWL)*

Four different flight performance ratings (altitude precision, speed precision, start turn, banking turn), from analyses of flight data (altitude, variation in altitude, air speed, variation in air speed, heading, attitude, start of turn and banking turn), were performed by simulator instructors. The mean of the four ratings formed a flight performance index. Furthermore, instructors rated (from protocols) the pilots' ability to control and report changes in the tactical situation (i.e. their TSA). This rating was named TSA. The ratings were made on a seven point rating scale.

Figure 8 presents the relationship between the pilots, flight performance and their tactical situational awareness. There was a positive relation between ratings of flight performance and handling of tactical information ($r = 0{\cdot}59, p < 0{\cdot}001$), as can be seen. Those pilots who performed well with respect to the flight task also performed well with respect to the information handling task and *vice versa.* If the pilots had given priority to one of the tasks (flying or information handling) under high information load, a negative relationship would have been expected.

Directly after each sortie the pilots answered a questionnaire. The questions formed a basis for diagnoses or analyses of the PMWL measures as well as the tasks performed. By means of iterative principal axis factor analysis the number of items has been reduced to 10 indices. The number of factors was determined by means of the scree test. This procedure provides a solution with the minimum number of factors accounting for the maximum amount of variance. The solution accounted for 61 percent of the total variance. Items with significant factor loadings formed the indices. The mean of the factor loadings was 0·66 (range 0·40–0·95). The reliabilities of the indices were estimated by means of Cronbach's alpha. Table 1 presents the indices, the number of markers or items and reliability values. The indices have an acceptable to high reliability except for motivation.

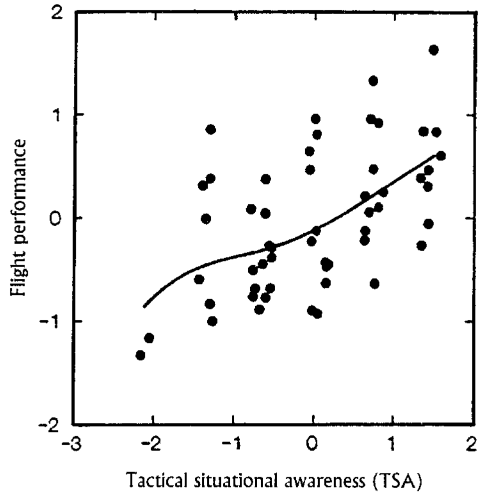


Figure 8. Instructor ratings of the pilots' flight performance as a function of instructor ratings of their Tactical Situational awareness (TSA). The curve has been smoothed by means of DWLS. The variables are Z-transformed.

Table 1. Ten indices of analysis.

Name	number of items	reliability values
Performance	3	0·87
Difficulty	3	0·73
Motivation	3	0·67
Mental capacity	4	0·76
Mental effort	4	0·86
Mental workload	4	0·90
Time stress	4	0·86
Saturation TSD	4	0·84
Complexity TSD	5	0·89
Balance flight-IH	4	0·77

None of the indices differentiated between the missions with tactical information presented on TSD, but nine out of ten differentiated between three of them and the training run ($p < 0.05$).

The pilots responded to BFRS (Roscoe 1987, Roscoe and Ellis 1990) and the six items of NASA-TLX (Hart and Staveland 1988) after the sorties and at four times during the sorties they responded to the three items of SWAT (Reid and Nygren 1988). The reliability estimates for BFRS, 'NASA-TLX' and 'SWAT' (Cronbach's alpha, based upon the covariances between the four runs) were satisfactory (0·82, 0·77 and 0·74, respectively) and in accordance with those found in other studies (Lysaght *et al.* 1989).

Table 2 gives the correlations (product moment) between the ten indices in table 1 and the three PMWL measures. Seventy-one percent of the correlations are significant ($p < 0.05$) and the structure of the matrix is consistent with expectations. For example, difficulty covaries with mental workload, mental effort, time stress, mental capacity and performance. The significant correlations between saturation and complexity of TSD *and* difficulty, mental workload, mental effort, performance, capacity and time stress indicate that the handling of tactical information on TSD dominated the situation.

Table 2. The correlations (product moment) between the 10 indices of table 2 and the three MWL measures.

	Perf.	Diff.	Mot.	Cap.	Eff.	MWL.	Time	Bal.	Sat.	Comp.	BFRS	'NTLX'	'SWAT'
Performance (Perf)	1·000												
Difficulty (Diff)	–0·432	1·000											
Motivation (Mot)	0·447	0·228	1·000										
Mental capacity (Cap)	0·393	–0·664	–0·093	1·000									
Mental effort (Eff)	–0·269	0·848	0·259	–0·706	1·000								
Mental workload (MWL)	–0·151	0·787	0·321	–0·664	0·908	1·000							
Time stress (Time)	–0·357	0·757	0·231	–0·730	0·850	0·811	1·000						
Balance flight-IH (Bal)	–0·101	–0·055	0·125	–0·049	0·047	0·063	–0·052	1·000					
Saturation TSD (Sat)	–0·488	0·420	0·165	–0·609	0·360	0·300	0·476	0·123	1·000				
Complexity TSD (Comp)	–0·362	0·506	0·135	–0·447	0·423	0·465	0·550	–0·054	0·432	1·000			
BFRS	–0·374	0·678	0·210	–0·798	0·675	0·697	0·726	0·083	0·504	0·409	1·000		
'NASA-TLX'	–0·427	0·840	0·191	–0·749	0·868	0·840	0·828	0·154	0·473	0·430	0·826	1·000	
'SWAT'	–0·185	0·686	0·078	–0·569	0·711	0·639	0·544	0·052	0·194	0·108	0·687	0·735	1·000

Significant correlations ($p < 0{\cdot}05$) are underlined.

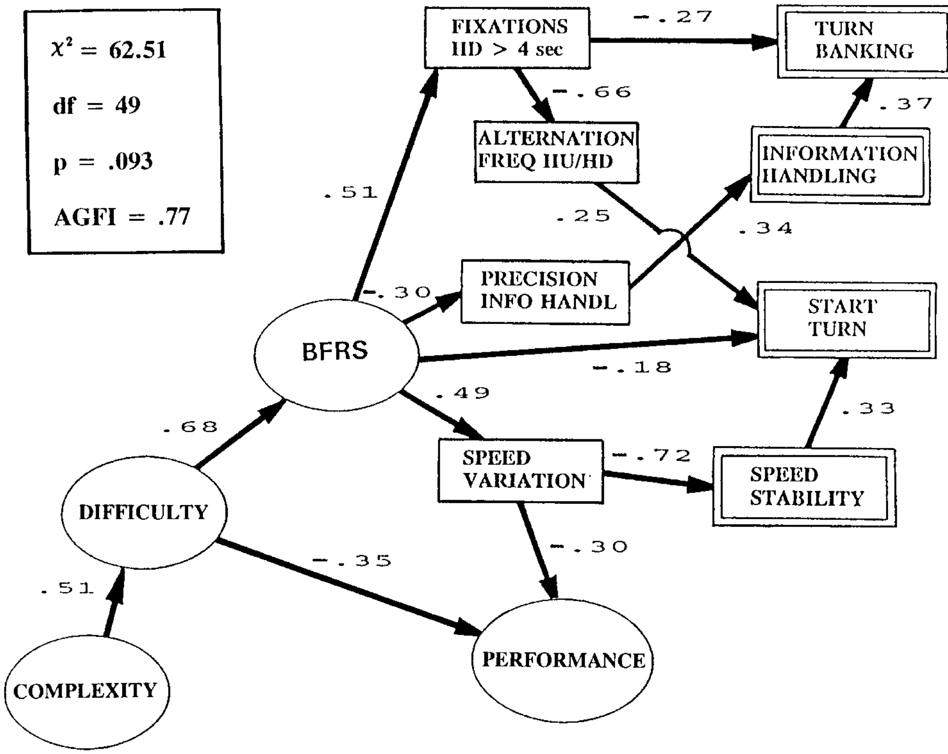


Figure 9. Structural LISREL model of the relationships between the indices, PMWL measures, objective registration and instructor ratings. Pilot ratings are marked by ellipses, objective registrations by simple rectangles and instructor ratings by double rectangles. Effects $>0{\cdot}24$ are significant ($p<0{\cdot}05$). Adjusted Goodness of Fit Index (AGFI) $=0{\cdot}77$.

The table shows that the correlations between the three PMWL measures (BFRS, 'NASA-TLX' and 'SWAT') are significant ($p<0{\cdot}001$) and in accordance with the results of other research (Lysaght *et al.* 1989). Even if the correlations are high (0·69–0·83), the proportions of unexplained variance are still considerable (31–53%). Factors that possibly contribute to this incongruity are differences in context and time of measurement (during and after the sorties, respectively).

The PMWL measures covaried with the indices difficulty, mental effort, mental workload, time stress, capacity, TSD saturation, TSD complexity and performance (i.e. eight out of the ten indices).

Mental workload increased as a function of difficulty (increasing TSD complexity). The increase in workload was compensated for by an increased mental effort; the increase in workload, sooner or later, resulted in decreased performance.

The paragraph above forms an embryo to model how different factors affect and are affected by PMWL. Causal models of this type can be tested statistically (Jöreskog and Sörbom 1984) by means of Structural Equation Modeling (LISREL VI). The model presented in figure 9 has been tested based upon the correlations between the indices of table 2, the PMWL measures, objective registrations and instructor ratings. The fit of the model is acceptable (Adjusted Goodness of Fit Index (AGFI) $=0{\cdot}77$). Ellipses denote pilot ratings (BFRS, complexity TSD, difficulty and performance). Simple rectangles denote objective registrations (critical eye fixations HD, alternation frequencies HU-HD, precision information handling and variation of speed). Double rectangles denote instructor ratings (turn banking, information handling, start turn and stability of speed).

The model has its starting point in the task related indices of TSD complexity and difficulty and its terminal point in different aspects of the pilots' performance. The PMWL forms an intervening process.

The perceived difficulty increased as a function of the complexity of the tactical information. The increasing difficulty resulted, sooner or later, in decreased performance. The difficulty of the task had a strong effect on PMWL, which is of specific interest.

An increased PMWL caused changes in the objective registrations. The number of critical eye fixations (HD) and the variation in speed increased, while the precision of information handling (TSA) and the alternation frequency HU-HD decreased as a function of PMWL.

Changes in the objective registrations affected the instructor ratings of PP. An increased number of critical eye fixations HD has a negative influence on the performance ratings of turn banking. A low alternation frequency (HU-HD) had a negative influence on start turn (it was delayed). Ratings of information handling (TSA) decreased as a function of the difference between presented and reported number of objects on TSD. The ratings of speed stability decreased as a function of an increased speed variation, while an increased speed variation caused a decreased pilot rating of performance.

Summarizing, mental workload was sensitive to changes in the difficulty levels of the task and PMWL predicted objective as well as subjective measures of PP. The model is the same, if we replace BFRS by 'NASA-TLX', even if the fit of the model in that case is somewhat lower.

Several of the mood dimensions rated before the runs correlated significantly with the indices of table 2 and the PMWL measures. For instance, the dimension hedonic tone (happy–sad) correlated significantly with the indices of performance and difficulty (0·60 and −0·33, respectively, $p < 0{\cdot}01$).

The covariation between mood and the PMWL measures is of specific interest. The correlations between hedonic tone (rated before the run) *and* 'SWAT', 'NASA-TLX' and BFRS were −0·35, −0·44 and −0·42, respectively ($p < 0{\cdot}01$). The more alert, active, relaxed and confident the pilot was before a run, the better was his ability to cope with the mental load of it.

3.4. *Psychophysiological measures of mental workload*

Subjective measures apart, we have used heart rate (*HR*) as a psychophysiological measure of PMWL. Sortie means, running means and variance have been calculated. The variance measure (S^2_{diff}) was based on the differences between successive *R-R* intervals. The frequency components of the ECG-signal have been calculated by means of Fast Fourier Transform (FFT).

The heart rate measures (mean and variance) have been correlated with the PMWL measures and the indices of table 2. Heart rate (mission means) increased as a function of the indices difficulty and TSD complexity ($r = 0{\cdot}37$ in both cases, $p < 0{\cdot}01$). Of the PMWL measures, only 'SWAT' correlated significantly with *HR* ($r = 0{\cdot}34$, $p < 0{\cdot}05$).

According to the literature (cf. Wilson and Eggemeier 1991), we could expect a decrease of *HR* variance as a function of mental workload. However, a significant *positive* relationship was found between the variance measure S^2_{diff} and TSD complexity ($r = 0{\cdot}37$, $p < 0{\cdot}01$). Thus, the variance increased as a function of mental load. However, we found a positive relation between mean *HR* and S^2_{diff} which may explain our result. Another possible explanation is that S^2_{diff} was an inadequate measure of heart rate variability.

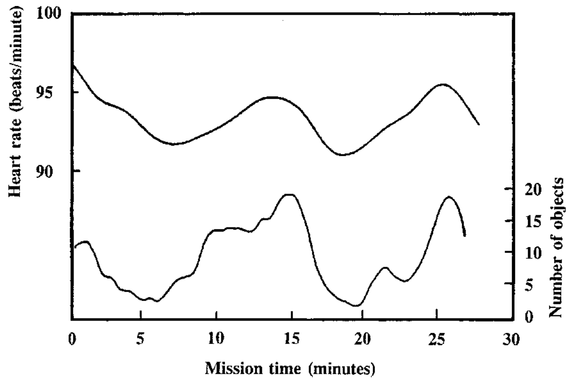


Figure 10. Heart rate (upper curve) and number of objects on TSD (lower curve) as a function of mission time for the 'expert' pilot A. The curves have been smoothed by means of DWLS.

Spectral analyses (FFT) of cardiac interval times have been performed. Power density spectra from periods (of 256 seconds) with high information load have been compared with periods with low information load. It was found that the power of the low frequency band (0·01–0·06 Hz) and the power of the frequency band around 0·10 Hz, decreased during periods of high PMWL as compared to periods of low PMWL. The decreased amplitude of the 0·10 Hz component is in agreement with the results of others (Aasman *et al.* 1987). The effects were more pronounced for the pilots who performed well.

Heart rate was registered continuously, unlike the subjective PMWL ratings. This makes an analysis of changes in psychophysiological activation as a function of changes in information load during the sorties possible. We found that *HR* (running means) covaried significantly with variations in information load over the sortie for those pilots who had above mean performance. The mean correlation between *HR* and information load for these nine pilots was 0·33 ($p =< 0{\cdot}01$). (This coefficient is an underestimate of the true covariation because it is based on unfiltered data and not on running means.) The correlation between the 18 pilots' rank order and the covariations (*HR*-information load) was $-0{\cdot}55$ ($p = 0{\cdot}019$). Thus, the diagnostic value of *HR* is affected by the pilots' skill level. Figure 10 presents the covariation between *HR* and the number of objects presented on TSD as a function of mission time for the 'expert' pilot A. The common variance between the curves is 55 percent.

4. Discussion

4.1. *Summary and conclusions*

The primary purpose of the present study was to: analyse the pilot's ability to perceive and use synthetic information, study the effects of information complexity on PMWL and PP, and analyse the structure of PMWL.

The importance of paying attention to human limitations for information processing has been clearly documented in the present results. Although the point is somewhat obvious to experts on man-machine problems, it has apparently been more or less neglected by designers of even modern aircraft, perhaps because modern computer technology offers such tempting possibilities to develop complex and, in principle, very informative displays.

Eighteen pilots performed 72 simulated low level–high speed missions. The complexity of the HDD information varied as a function of the tactical situation. Flight data (altitude, air speed, heading) were registered continuously. The pilots' eye movements were video taped and their psychophysiological activation in terms of heart rate was saved on disks. During and after the sorties, the pilots rated their mental workload according to the psychological content of three scales (BFRS, SWAT, NASA-TLX) and a questionnaire of 60 assorted items.

It was found that even a moderate complexity of information interfered with the flight task. The pilots' altitude and variation in altitude increased and their corrections of altitude errors were delayed. Durations and frequencies of eye fixations (HU-HD) changed as a function of information load. When the information load increased, the pilots controlled the surrounding world less often and the inspection times HU became shorter and shorter. Thus, the conditions of flying low level–high speed with high precision deteriorated.

The error of the pilots handling of tactical information (i.e. their tactical situational awareness) increased, on average, by one object (defence areas or aircraft), when the complexity was increased by five objects. Over- and underestimations leveled each other out up to about ten objects. When the number of objects increased further, the underestimations began to dominate, and the ability of the pilots to detect changes in the tactical situation deteriorated drastically. The interindividual variance also increased as a function of the complexity of TSD. One less skillful pilot (B) began to underestimate the number of objects, i.e. miss critical objects or threats, at a complexity of about six objects. The corresponding value for an 'expert' pilot (A) was 15 objects.

Pilots who performed well with respect to the flight task also performed well with respect to the information handling and *vice versa*. If the pilots had given priority to one of the tasks (flying or information handling) under high information load, a negative relationship would have been expected. When the information load increased to more than 8–10 objects on TSD, the pilots could no longer integrate the two tasks (information handling and flying) and they succeeded in neither of them. The results also demonstrated interindividual differences in skill. The skilled pilots seemed to use more effective strategies to integrate information (HU and HD) which promoted their flight performance and situational awareness; the pilots who performed well perceived less load from the TSD information, their fixation times HD were shorter and their alternation rates (HU-HD) were higher.

The results are in accordance with scientific literature on human information processing. During almost four decades, one has been able to read about 'the magical number seven' and its significance for attention, short term memory, discrimination and decision making (Miller 1956, cf. Baddeley 1994). Humans have severe limitations in what they can receive, process and remember. We saw how flight performance was affected by the information load: The pilot increased his/her altitude with the complexity of TSD. This could be an accurate way to reduce workload—and perceived risk level—in a simulated situation and in the air as well, under peacetime conditions. In war, however, the pilot does not have this option; increased altitude is always connected with exposure to enemy air defence threats and therefore an increased risk for the pilot. The optimum performance is probably a function of pilot skill, flight task difficulty and information load. A pilot can probably process more complex information simultaneously with an easier flight task.

It was found from a structural model, that mental workload was affected by the complexity of mission and that workload affected different aspects of performance. The complexity of the tactical information on TSD had a strong impact on mental workload, which affected eye point of gaze, awareness of the tactical situation and speed precision. The relationship between PMWL and performance was nonlinear.

High correlations, observed here, are to be expected due to similarity of item content. The fact that these items were responded to in close temporal contiguity may have caused some further statistical overlap. However, correlations rarely increase very strongly due to the time factor alone; content similarity is a much more important factor.

It is interesting to note that several mood dimensions rated before the sorties correlated significantly with the indices of performance and difficulty as well as with the PMWL measures. The more alert, active, relaxed and confident the pilot was before a run, the better his ability to cope with the mental load. When it concerns performance, the predictive power of mood dimensions as hedonic tone, activation, tension and control, has been demonstrated in several studies (Angelborg-Thanderz 1990, Svensson *et al.* 1988, 1993).

Heart rate (sortie means and variance) correlated positively with PMWL and perceived complexity of mission. Heart rate (running means) covaried with variations in information load over the sorties for those pilots who performed well. From spectral analyses of cardiac interval times it was found that the amplitude of the 0·10 Hz component decreased during high levels, as compared to low levels, of information load. There were potential artifacts (e.g. long trends, respiration rate) and the results may have been diluted.

As for the measurement of PMWL the psychological content of the present procedures was in accordance with alternatives available from earlier work in the USA and the UK. It must be stressed, however, that the measurements of PMWL *per se* is favoured and that its antecedents (e.g. task properties such as risk and difficulty) and their consequences (e.g. perceived success) should be measured separately, favouring conceptual clarity in the interpretations of results. This is, in fact, the approach taken elsewhere (Svensson *et al.* 1993) where it was found to be quite feasible to model the interplay between PMWL and its antecedents and consequences. The present study provides some further support for this strategy.

The prescribed scale formats and rating procedures for SWAT and NASA-TLX have not been used in this study. However, we have reason to believe that this has not affected the results to any great extent; in a recent study of pilot performance and mental workload we have compared a weighted and an unweighted variant of NASA-TLX. The correlation between the two variants was 0·97 ($p < 0{\cdot}001$). The psychometric properties of the different procedures will be discussed elsewhere.

The methods used here to measure pilot reactions are rather complex and time-consuming. In further research, and in practical applications, they should be simplified.

Dynamic situations need dynamic measures. Most subjective measures of pilot mental workload (as well as pilot performance and situational awareness) are, more or less, static (i.e. reflect means over periods of minutes or more). We believe that dynamic subjective measures will increase the diagnostic power of workload and performance measures. In a cooperative project within the European Community (VINTHEC, Visual Interaction and Human Effectiveness in the Cockpit) dynamic subjective measures of pilot performance, mental workload and situational awareness are developed. The performance measures are based upon a pilot function model developed by Angelborg-Thanderz (1989, 1990).

Finally, a few more words on information overflow. The present findings are based on a small sample, but they are quite clear and in accordance with previous work on information overload. Modern technology makes it feasible to provide the pilot with a wealth of information. Certainly, the limit of performance is set by the limitations of the human operator and not by technological possibilities. In this situation we feel that a realistic conception of the human limitations and efficacious psychological models of the pilot and his/her performance are of fundamental importance in the design of combat aircraft systems.

References

Aasman, J., Mulder, G. and Mulder, L. J. M. 1987, Operator effort and the measurement of heart-rate variability, *Human Factors*, **29**, 161–170.

Angelborg-Thanderz, M. 1989, *Assessing pilot performance in training simulators. A structural analysis.* In 1989 Spring Convention-flight simulation: Assessing the benefits and economics, The Royal Aeronautical Society, London.

Angelborg-Thanderz, M. 1990, *Military flight training at a reasonable price and risk*, Economics Research Institute, Stockholm School of Economics and FOA report C 50083-5.1. (report in Swedish, summary in English).

Baddely, A. 1994, The magical number seven: still magic after all these years? *Psychological Review*, **101**, 353–356.

Casali, J. G. and Wierwille, W. W. 1983, A comparison of rating scale, secondary task, physiological, and primary task workload estimation techniques in a simulated flight task emphasizing communication load, *Human Factors*, **25**, 623–642.

Chase W. G. and Ericsson, K. A. 1981, Skilled memory. In J. R. Anderson (ed.), *Cognitive skills and their acquisitions. The sixteenth annual Carnegie Mellon symposium on cognition* (Erlbaum, Hillsdale), 141–189.

Cooper, G. E. and Harper, R., Jr. 1969, *The use of pilot rating in the evaluation of aircraft handling qualities*, NASA, Moffett Field, CA, Report No. NASA TN-D-5153.

Easterbrook, J. A. 1959, The effect of emotion on cue utilization and the organization of behavior, *Psychological Review*, **66**, 183–201.

Eggemeier, F. T. and Wilson, G. F. 1991, Performance-based and subjective assessment of workload in multi-task environments. In D. L. Damos (ed.), *Multiple task performance* (Taylor & Francis, London), 217–278.

Eustase, H. F. 1977, *ICH, The International Countermeasures Handbook* (EW Communications, Palo Alto), 404–405.

Gilson, R. D. 1995, *Situational awareness* (special issue), *Human Factors*, **37**.

Goebel, G. M. 1983, Got a minute? A further look at 'Vigilance and distraction', *TAC ATTACK*, **23**, 18–21.

Gopher, D. and Donchin, E. 1986, Workload—an examination of the concept. In K. R. Boff, L. Kaufman and J. P. Thomas (eds), *Handbook of perception and human performance. Vol. II.* (Wiley, New York), 41-1–41-49.

Gunston, B. 1986, *Jane's Aerospace Dictionary* (Jane's, London).

Hart, S. G. and Staveland, L. E. 1988, Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (eds), *Human Mental Workload* (Elsevier, Amsterdam), 139–183.

Jensen, S. E. 1994, Developing a flight workload profile using Continuous Subjective Assessment of Workload (C-SAW). In R. Fuller, N. Johnston and N. McDonald (eds), *Human factors in aviation operations* (Averbury Aviation, Aldershot), 307–312.

Jöreskog, K. G. and Sörbom, D. 1984, *LISREL VI. Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods*, Department of Statistics, University of Uppsala, Sweden.

Klein, G. A. 1993, *Naturalistic decision making: implications for design*, Crew Systems Ergonomics, Information Analysis Center, Wright-Patterson AFB, Reports No. CSERIAC SOAR 93-1.

Lysaght, R. J., Hill, S. G., Dick, A. O., Plamondon, B. D., Linton, P. M., Wierwille, W. W., Zaklad, A. L., Bittner, Jr., A. C. and Wherry, R. J. 1989, *Operator workload: Comprehensive review and evaluation of operator workload methodologies*, United States Army Research Institute for the Behavioral Sciences, Technical Report 851, MDA 903-86-C-0384.

Miller, G. A., 1956, The magical number seven, plus or minus two, *Psychological Review*, **63**, 81–97.

Moray, N. 1986, Monitoring behavior and supervisory control. In K. R. Boff, L. Kaufman and J. P. Thomas (eds), *Handbook of Perception and Human Performance. Vol. II. Cognitive Processes and Performance* (Wiley, New York), 40-1–40-51.

Muckler, F. A. and Seven, S. A. 1991, Selecting performance measures: 'Objective' versus 'subjective' measurement, *Human Factors*, **34**, 441–455.

O'Donnell, R. D. and Eggemeier, F. T. 1986, Workload assessment methodology. In K. R. Boff, L. Kaufman and J. P. Thomas (eds), *Handbook of perception and human performance. Vol. II* (Wiley, New York), 42-1–42-49.

van Raaij, W. F. 1988, Information processing and decision making. Cognitive aspects of economic behavior. In W. F. van Raaij, G. M. van Veldhoven and K. E. Wärneryd (eds), *Handbook of Economic Psychology* (Kluwer Academic, Dordrecht), 74–106.

Reid, G. B. and Nygren, T. E. 1988, The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock and N. Meshkati (eds), *Human mental workload* (Elsevier, Amsterdam), 184–218.

Roscoe, A. H., 1987, In-flight assessment of workload using pilot ratings and heart rate. In A. H. Roscoe (ed.), *The practical assessment of pilot workload*, AGARDograph No. 282.

Roscoe, A. H. and Ellis, G. A. 1990, *A subjective rating scale for assessing pilot workload in flight: A decade of practical use*, Royal Aerospace Establishment, Technical Report 90019.

Simon, H. A. 1990, Invariants of human behavior, *Annual Review of Psychology*, **41**, 1–19.

Sjöberg, L., Svensson, E. and Persson, L. O. 1979, The measurement of mood, *Scandinavian Journal of Psychology*, **20**, 1–18.

Svenson, O. and Benson, L. 1991, *Framing and time pressure in decision making*, Psychological Research Bulletin XXXI: 1, Department of Psychology, University of Lund, Sweden.

Svenson, O. and Edland, A. 1987. Change of preferences under time pressure: choices and judgments, *Scandinavian Journal of Psychology*, **28**, 322–330.

Svensson, E., Angelborg-Thanderz, M., Sjöberg, L. and Gillberg, M. 1988, Military flight experience and sympatho-adrenal activity, *Aviation, Space, and Environmental Medicine*, **59**, 411–416.

Svensson, E., Angelborg-Thanderz, M. and Sjöberg, L. 1993, Mission challenge, mental workload and performance in military aviation, *Aviation, Space, and Environmental Medicine*, **64**, 985–991.

Thanderz, M. 1973, *A study of pilot training on a modern weapon system*, Institute of Aviation Medicine, Flight Test Centre, Air Materiel Department, Defence Materiel Administration, Sweden, Report FMI-73-10.

Thanderz, M. 1982, *Assessing pilot performance and mental workload in training simulators*, Flight simulation, Aviation medicine, and Avionics system groups Two Day International Conference, The Royal Aeronautical Society, London.

Way, T. C., Martin, R. L., Gilmour, J. D., Hornsby, M. E. and Edwards, R. E. 1987, *Multicrew pictorial format display evaluation*, Flight Dynamics Laboratory, Air Force Wright Aeronautical Laboratories, Air Force Systems Command, Wright-Patterson Air Force Base, AFWAL-TR-87-3047.

Wickens, D. S. 1987, Information processing, decision-making, and cognition. In G. Salvendy (ed.), *Handbook of Human Factors* (Wiley, New York), 72–107.

Wickens, C. D. and Yeh, Y. Y. 1983, The dissociation between subjective workload and performance: a multiple resource approach, *Proceedings of the Twenty-Seventh Annual Meeting of the Human Factors Society*, **27**, 244–248.

Wierwille, W. W. and Casali, J. G. 1983, A validated rating scale for global mental workload measurement applications, *Proceedings of the Human Factors Society*, **27**, 129–133.

Wierwille, W. W. and Connor, S. A. 1983, Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator, *Human Factors*, **25**, 1–16.

Wilson, G. F. and Eggemeier, F. T. 1991, Psychophysiological assessment of workload in multi-task environments. In D. L. Damos (ed.), *Multiple task performance* (Taylor & Francis, London), 329–360.